

Research Article

Experiments on Auditory-Visual Perception of Sentences by Users of Unilateral, Bimodal, and Bilateral Cochlear Implants

Michael F. Dorman,^a Julie Liss,^b Shuai Wang,^b Visar Berisha,^b
Cimarron Ludwig,^b and Sarah Cook Natale^b

Purpose: Five experiments probed auditory-visual (AV) understanding of sentences by users of cochlear implants (CIs).

Method: Sentence material was presented in auditory (A), visual (V), and AV test conditions to listeners with normal hearing and CI users.

Results: (a) Most CI users report that most of the time, they have access to both A and V information when listening to speech. (b) CI users did not achieve better scores on a task of speechreading than did listeners with normal hearing. (c) Sentences that are easy to speechread provided 12 percentage points more gain to speech understanding than did sentences that were difficult. (d) Ease of speechreading for sentences is related to phrase familiarity.

(e) Users of bimodal CIs benefit from low-frequency acoustic hearing even when V cues are available, and a second CI adds to the benefit of a single CI when V cues are available. (f) V information facilitates lexical segmentation by improving the recognition of the number of syllables produced and the relative strength of these syllables.

Conclusions: Our data are consistent with the view that V information improves CI users' ability to identify syllables in the acoustic stream and to recognize their relative juxtaposed strengths. Enhanced syllable resolution allows better identification of word onsets, which, when combined with place-of-articulation information from visible consonants, improves lexical access.

For most neuroscientists, speech perception is inherently multimodal—that is, the brain normally integrates information from the auditory system, the visual system, even the tactile system (Gick & Derrick, 2009), to segment the continuous auditory signal and to constrain and access lexical representations (e.g., Campbell, 2008; Peelle & Sommers, 2015; Rosenblum, 2005; Summerfield, 1987). From this point of view, speech perception is not primarily auditory. Moreover, visual information about speech is not information “added onto” the information provided by the auditory signal; rather, cortical systems underlying speech perception are inherently sensitive to the information provided by multiple modalities. Indeed, from one point of view, speech perception is amodal and involves the extraction of common higher-order information from both auditory and visual signals (Rosenblum, 2008).

In the United States, with very few exceptions, the assessment of speech understanding by users of cochlear implants (CIs) is conducted in auditory-only test conditions. This is reasonable because the speech-understanding score is used most commonly as a surrogate for the degree to which the CI has successfully replaced the damaged peripheral auditory system. For this reason, there is a very large literature on speech understanding by CI users in auditory-only test conditions but a much smaller literature on situations in which both auditory (A) and visual (V) sources of information are available.

To explore speech understanding by CI users in auditory-visual (AV) test conditions, we conducted five experiments. These experiments evolved from the findings of a survey of CI users in which they were asked about the environments in which they listened to speech. The results indicated that most CI users, most of the time, have access to both A and V information when listening to speech. In subsequent experiments we explored (a) the ability of listeners with normal hearing and of CI users to speechread sentence material designed for AV testing; (b) the benefit to AV speech understanding of sentences that are easy and difficult to speechread; (c) the factors responsible for making sentences easy or difficult to speechread; (d) the value to speech

^aDepartment of Speech and Hearing Science, Arizona State University, Tempe

^bArizona State University, Tempe

Correspondence to Michael Dorman: mdorman@asu.edu

Editor: Nancy Tye-Murray

Associate Editor: Richard Dowell

Received September 8, 2015

Revision received February 12, 2016

Accepted April 4, 2016

DOI: 10.1044/2016_JSLHR-H-15-0312

Disclosure: The authors have declared that no competing interests existed at the time of publication.

understanding of low-frequency acoustic hearing in the contralateral ear to a CI and of a second CI (bilateral CIs) when V information is available; and (e) the role of V information in providing information about lexical boundaries.

Experiment 1: The Environments in Which CI Users Listen to Speech

As noted, there is a very large literature on speech understanding by CI users in auditory-only test environments. If CI users commonly listen in auditory-only environments—for example, on the telephone or to the radio—then these data are widely relevant. However, if listening in an auditory-only environment is not common, then the data will not provide a generally accurate answer to the question asked by every patient: “How well will I understand speech with a cochlear implant?”

It is reasonable to suppose that CI users, like other listeners with impaired hearing (Humes, 1991), will seek out environments in which they have both A and V input for speech understanding. However, reasonable suppositions do not rise to the level of data. For this reason, we asked CI users to report, via an online survey, the environments in which they most commonly listen to speech.

Method

Subjects

Following approval by the Institutional Review Board at Arizona State University, an invitation to participate in an anonymous survey was sent via e-mail to 413 CI users who had previously indicated an interest in participating in research. Responses were obtained via SurveyMonkey (<https://www.surveymonkey.com/>) from 131 of them—a 31.7% response rate.

Demographic data collected from the survey indicated that the subjects ranged in age from 21 to 81 years. Most of the subjects (86%) were between ages 51 and 81 years. Sixty-one percent of the subjects were bilateral CI users and 39% had a single CI. Most of the unilateral users (81%) had used their implants for more than 5 years, and most of the bilateral users (86%) had used two implants for 3 or more years.

Questionnaire

The questions related to listening with a CI were the following:

1. How much do you interact with other people in a typical day? (Response possibilities: No interaction; Some interaction; Regular interaction; Extensive interaction.)
2. Which sources of speech do you encounter on most days? (Response possibilities: Radio; Television when I cannot see the speaker; Television when I can see the speaker; Another person who I can see; Telephone conversations. Multiple answers were possible for this item.)

3. The most common sources of speech I hear come from _____. (Response possibilities: Radio; Television when I cannot see the speaker; Television when I can see the speaker; Another person who I can see; Telephone conversations.)
4. When having a conversation, how often can you see the speaker's face? (Response possibilities: Almost never; Less than half of the time; Most of the time; Almost all of the time.)

Results and Discussion

The results are shown in Figure 1. Our demographic data, combined with the data in Figures 1a and 1b, paint a picture of an older, socially active group of experienced CI users who encounter multiple sources of speech input each day. Nearly three quarters of the users could understand speech sufficiently well with their CI alone that they used the telephone on a regular basis (for a review of telephone use among CI users, see Clinkard et al., 2011). Nearly half indicated that they regularly listened to the radio. These outcomes are consistent with many reports showing very high levels of sentence recognition ($\geq 80\%$ correct for the best-aided condition) in quiet for CI users when tested in A test conditions with “everyday” sentences (e.g., Gifford, Shallop, & Peterson, 2008; Massa & Ruckenstein, 2014; Wanna et al., 2014).

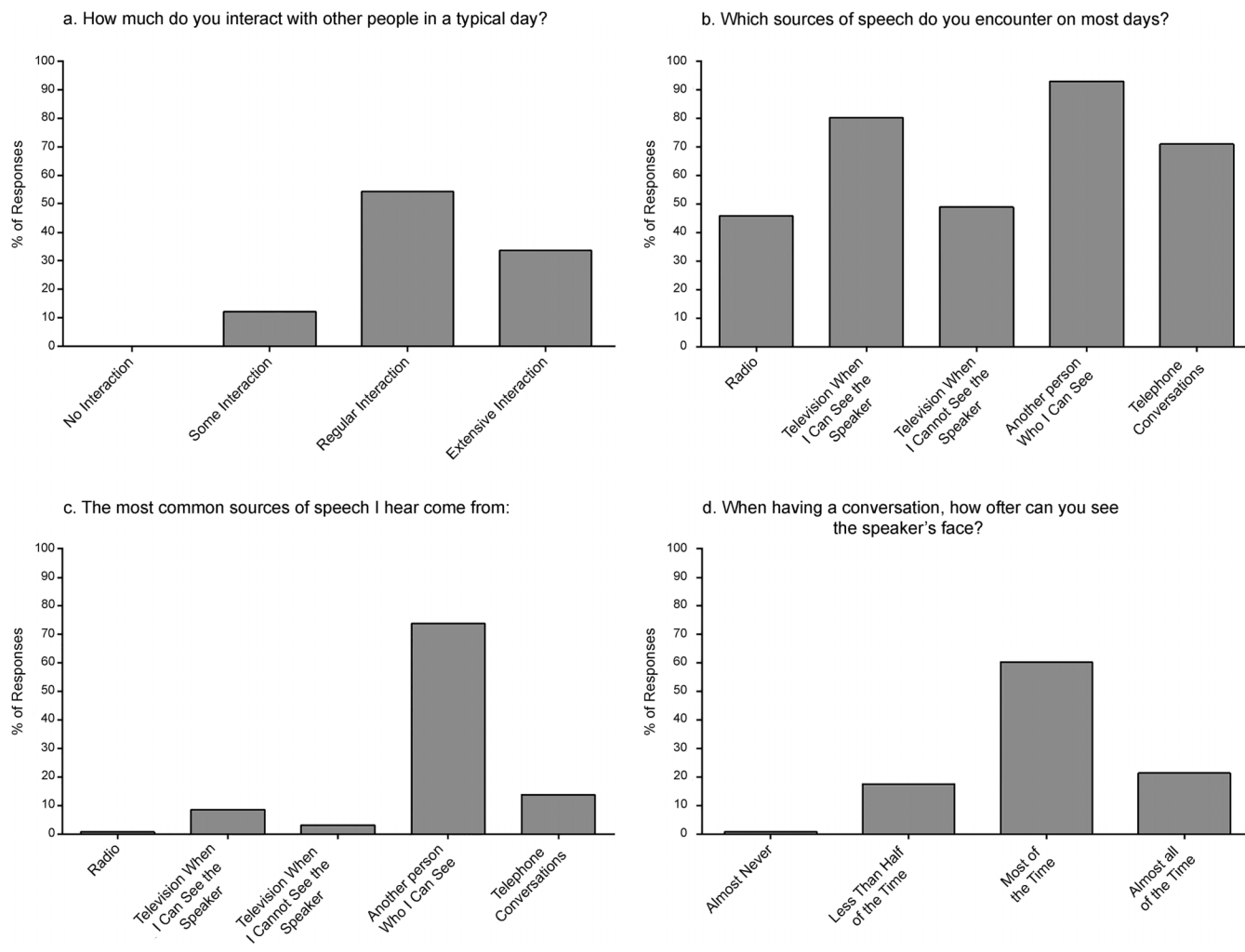
Inspection of Figures 1c and 1d reveals that the most common source of speech encountered on a daily basis was speech from another person and that most or almost all of the time, the listener could see the speaker's face. Thus, as we supposed, a combination of A and V information is the basis for speech perception for most CI users most of the time.

Although the subjective data from our questionnaire appear unambiguous, individuals can interpret descriptors such as “some,” “most,” “regular,” and “extensive” differently (see, e.g., Wikman, 2006). An objective study using CI-based data-logging systems and eye-tracking hardware and software would add to our knowledge of listening environments and the time that CI users spend in those environments.

Experiment 2a: Speechreading Abilities of Listeners With Normal Hearing and CI Users

The results of Experiment 1 suggest that, if we are to probe speech understanding in the most common listening environments, it would be useful to have tests of speech understanding that include both A and V information (e.g., Bergeson, Pisoni, Reese, & Kirk, 2003; Kirk et al., 2012; Robbins, Renshaw, & Osberger, 1995; Tye-Murray et al., 2008; Tyler, Preece, & Lowder, 1983; van Dijk et al., 1999). To that end, we have created sentence lists for AV testing—the AzAV sentences—that were developed using the methodology used by Spahr et al. in the creation of the AzBio sentences for adults (2012) and for children (2014) and share

Figure 1. Survey results for questions relating to listening environments encountered by CI users.



with those two tests the property of equal list intelligibility. In the following, we briefly describe the new material.

The AzAV sentences are a rerecording of an AV test created by Macleod and Summerfield (1987, 1990). These materials were recorded in British English and had equal across-lists auditory intelligibility for 10 lists and equal gain from visual information. To create the AzAV materials, the same sentences were recorded in American English by a single female speaker (Cook, Sobota, & Dorman, 2014). The sentence corpus contains 10 lists of sentences. Each list contains 15 sentences. Each sentence has the same “noun phrase–verb phrase–noun phrase” structure. List equivalence in an auditory-only condition and equal gain from vision were maintained in the new recording of the Macleod and Summerfield (1987) sentences. This was established by testing 10 young listeners with normal hearing using noise-vocoded signals. This outcome is of some interest, given the very different speakers and dialects of English—that is, a male speaker of British English and a female speaker of American English.

Previous Studies Using AV Test Material With CI Users

The benefits of adding V information to A information in speech recognition for CI users have been described for material at the consonant level (e.g., Desai, Stickney, & Zeng, 2008; Schorr, Fox, van Wassenhove, & Knudsen, 2005), word level (e.g., Gray, Quinn, Court, Vanat, & Baguley, 1995; Kaiser, Kirk, Lachs, & Pisoni, 2003; Rouger, Fraysse, Deguine, & Barone, 2008), and sentence level (e.g., Altieri, Pisoni, & Townsend, 2011; Bergeson et al., 2003; Most, Rothem, & Luntz, 2009; van Dijk et al., 1999). For this article, the most relevant data are those from tests of sentence recognition. This literature suggests that AV gain is related to the speechreading difficulty of the material—that is, the easier the material is to speechread, the greater the gain in an AV condition relative to an A condition. For that reason, if AV sentences, like those in the AzAV corpus, are to be used in clinical testing, then it is necessary to first determine the relative difficulty of the material in terms

of speechreading. If this is known, then it is possible to view the outcome of AV testing as a liberal or conservative estimate of the AV speech-understanding ability of the subject.

In order to assess and scale the speechreading difficulty of sentences, Kopra, Kopra, Abrahamson, and Dunlop (1986) created sentences that varied in speechreading difficulty. The Kopra corpus consists of 12 lists of sentences that are ranked from the highest speechreading index (95% correct) to the lowest (1% correct). Each list consists of 25 sentences.

For this study, sentences from the Kopra corpus were recorded in AV format. Lists 1, 4, 8, and 12 spanned the speechreading difficulty continuum from easy to difficult. Listeners were presented the Kopra sentence material as well as the AzAV sentence material. In Experiment 2a we (a) established the relative speechreading difficulty of the AzAV material compared to the Kopra sentences and (b) determined the speechreading ability of listeners with normal hearing versus CI users. In Experiment 2b we assessed, for CI users, the gain to A speech intelligibility when V information is easy or difficult to speechread.

Method

Subjects

Sixteen women (age range = 21–34 years; mean age = 26 years) with normal hearing (NH) were tested with Kopra Lists 1 and 12 and the AzAV sentences (two lists) in a vision-only condition. An additional eight women with NH in a similar age range viewed Kopra Lists 4 and 8. This condition was added at the end of the experiment to determine whether there was an orderly progression of scores from the most easy to most difficult material to speechread. Nineteen adult CI users with postlingual deafness (nine men, 10 women; age range = 21–83 years; mean age = 65 years) viewed the AzAV sentences (two lists each). All subjects had normal or corrected-to-normal vision.

Test Material

Lists 1 (easiest to speechread), 4, 8, and 12 (most difficult to speechread) were selected from the Kopra sentences. A young female English speaker, who previously had recorded the AzAV sentences, recorded AV clips of these four lists. The speaker's head and torso were visible in the AV recordings.

Procedure

The subjects sat in a sound-treated booth with a loudspeaker and a 21-in. video monitor at a distance of 12 in. Sentences were presented in either AV or V test conditions. Subjects typed their answers using a keyboard.

Results and Discussion

The speechreading scores for the Kopra sentences and the AzAV sentences by NH subjects, and for the AzAV sentences by the CI users, are shown in Figure 2. The mean speechreading score for List 1 was 55% correct (standard

error of mean [SEM] = 3); for List 4, 46% correct ($SEM = 4$); for List 8, 30% correct ($SEM = 5$); and for List 12, 7% correct ($SEM = 2$). The mean speechreading score for the AzAV sentences by NH listeners was 7% correct ($SEM = 2$). The mean speechreading score for the AzAV sentences by CI users was 11% correct ($SEM = 2$).

For the NH subjects, a one-way analysis of variance revealed a main effect of list for the four Kopra lists, $F(3, 44) = 53.03$, $p < .0001$. Sidak's multiple-comparisons test revealed that all mean scores were significantly different except for Lists 1 and 4.

For these subjects, the mean scores on Kopra List 12 and the AzAV sentences did not differ, $t(15) = 0.43$, $p = .67$. For that reason, the AzAV lists should be classified as difficult to speechread.

The scores for NH subjects and CI users did not differ for the AzAV sentences, $t(32) = 0.89$, $p = .39$. Thus, for this material, CI users are not better speechreaders than NH listeners.

Experiment 2b: Gain in Intelligibility as a Function of Speechreading Difficulty

Method

Subjects

Ten CI users with postlingual deafness (six women, four men; age range = 21–87 years; mean age = 64 years) participated in this experiment. These subjects were a subset of the CI users who participated in Experiment 2a.

Procedure

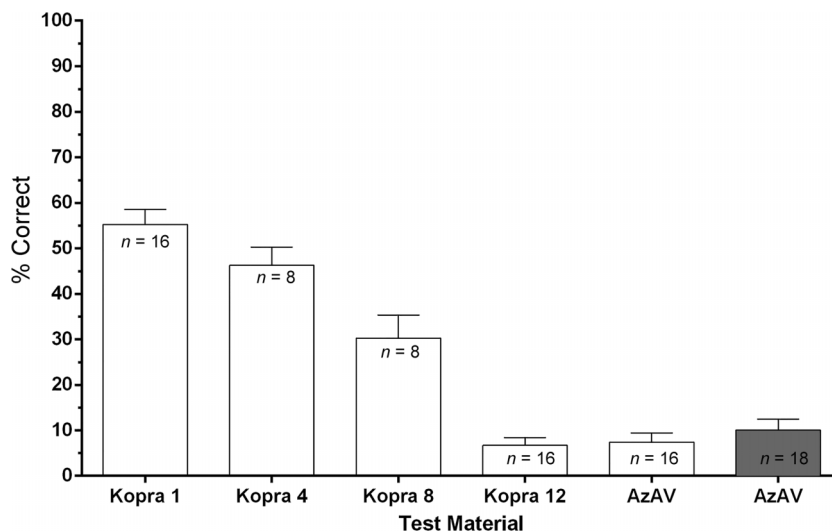
Lists 1 (easy) and 12 (difficult) from the Kopra sentences were used in both A and AV test environments. To minimize practice effects, the A tests and AV tests using the easy and difficult lists were performed on separate days. In addition, the list order was randomized. The target material was presented at 60 dB SPL. A small set of sentences was used to estimate the level of multi-talker babble necessary to drive each subject's A performance to near 40% correct in the easy condition. This speech-to-babble level, ranging from +10 to +3 dB, was then used with the test material. The test environment was the same as for the speechreading tests.

Results

The A and AV scores for the easy and difficult lists are plotted in Figure 3 for the CI users. The mean A score for the difficult list was 29% correct ($SEM = 3$). The mean AV score was 61% correct ($SEM = 5$). The gain from adding V information was 32 percentage points. The mean A score for the easy list was 41% correct ($SEM = 3$). The mean AV score was 85% correct ($SEM = 4$). The gain from adding V information was 44 percentage points.

A within-subject two-way analysis of variance revealed that test mode (A or AV), speechreading difficulty (easy or difficult), and the Test Mode \times Speechreading Difficulty

Figure 2. Percent correct word recognition in sentences as a function of type of test material in a vision-only condition. Sample size is indicated on each histogram. Open histograms = performance of listeners with normal hearing; gray histogram = performance of CI users; error bars = ± 1 SEM.

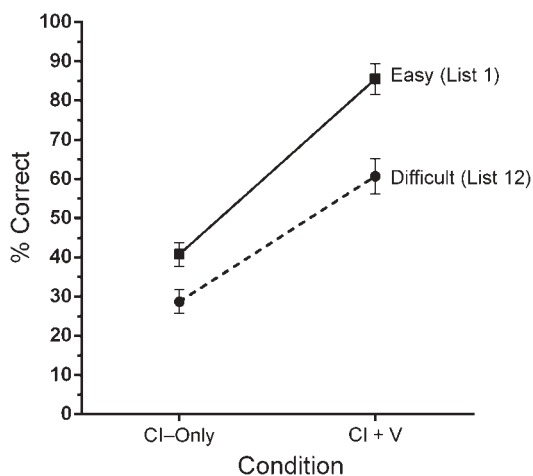


interaction all had significant effects on the scores—test mode: $F(1, 18) = 64.40, p < .0001$; speechreading difficulty: $F(1, 18) = 72.81, p < .0001$; interaction: $F(1, 18) = 11.88, p = .0029$. The significance of the interaction reflected the outcome that the sentences that were easier to speechread provided a larger visual benefit. The easy list provided 12 percentage points more visual benefit than the difficult list.

General Discussion

The results for the Kopra sentences validated the original claim by Kopra et al. (1986) that the lists of sentences were ordered from relatively easy to relatively

Figure 3. Percent correct word recognition as a function of test condition. The parameter is the speechreading difficulty of the material. CI = cochlear implant; V = vision; error bars = ± 1 SEM.



difficult on the basis of speechreading. We found that material that is easier to speechread provides more V gain than material that is difficult to speechread. This outcome is consistent with inferences from outcomes of other studies (e.g., Bergeson et al., 2003; Macleod & Summerfield, 1987; van Dijk et al., 1999).

For our subjects the mean scores on the most difficult Kopra list and the AzAV sentences did not differ. Thus, the AzAV lists should be classified as difficult to speechread. For that reason, the overall AV scores obtained with this material should be viewed as a conservative estimate of the potential AV scores obtained by CI users.

In this experiment CI users did not achieve higher speechreading scores than NH subjects. This is of interest because the NH subjects had almost no experience in the vision-only speechreading conditions, and presumably the CI users had extensive experience (for a review of speechreading by NH listeners and CI users in a variety of test environments, see Strelnikov, Rouger, Barone, & Deguine 2009).

Experiment 3: Phrase Familiarity and the Number of Visible Consonants as Factors Determining Speechreading Difficulty for Sentences

The results of Experiment 2 revealed that there can be large differences in speechreading difficulty among sentence lists. The aim of Experiment 3 was to investigate factors that determine speechreading difficulty.

Previous studies have demonstrated that number of visually perceptible segmental distinctions, lexical density, and word frequency can influence speechreading accuracy (e.g., Auer, 2002, 2009). In this experiment we applied

data-mining techniques (see Method) to explore, for Kopra Lists 1, 4, 8, and 12, the contributions to speechreading accuracy of low-level information—that is, the number of visible consonants (Summerfield, 1985)—and high-level information—that is, the degree to which word sequences, or phrases, are common in English.

Method

Phoneme Counts and Analyses

Text-mining processes that extracted words from the text of sentences in Lists 1, 4, 8, and 12 were done through the open-source Natural Language Tool Kit (Bird, Klein, & Loper, 2009). The Carnegie Mellon University Pronouncing Dictionary (<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>), which has over 125,000 words, was then used to decompose each word into a sequence of phonemes. Counts were made of all phonemes in each of the four lists.

Estimates of Phrase Familiarity

To estimate the familiarity of sentences, the numbers of occurrences of phrases (Google Ngrams) in approximately 5 million books published in the United States between 1986 and 2008 were retrieved through Google Books (Michel et al., 2011). To automate and standardize the process across all lists of Kopra sentences, each sentence was decomposed into a sequence of 3-grams (three-word phrases). For example, “Do you want a cookie?” was decomposed into “Do you want,” “you want a,” and “want a cookie.” The occurrence of each 3-gram in the 3-Gram American English database with smoothing factor 3 (Michel et al., 2011) from year 1986 to year 2008 was retrieved through a web crawler and the Natural Language Tool Kit platform. The mean, median, maximum, and minimum occurrence of phrases in each sentence were retrieved. In order to compare the occurrences across lists, the average of each metric on 25 sentences was computed.

Results and Discussion

All consonants were counted as a function of the list number. The result of this analysis for the visible consonants /m/, /b/, /p/, /w/, /s/, /ʃ/, /θ/, /ð/, /f/, and /v/ (Summerfield, 1985) is shown in Figure 4a. The overall count as a function of list number does not match the changes in speechreading accuracy shown in Figure 2 and reproduced in Figure 4b. However, the count of /w/ as a function of list number did show a trend similar to that of speechreading accuracy—that is, 18, 17, 11, and nine occurrences in Lists 1, 4, 8, and 12—and may have played a role in overall speechreading accuracy (for a model of the interaction of phonetic detail and higher-order constraints for *auditory* signals, in which small changes in phonetic detail can lead to large changes in speech understanding, see Boothroyd & Nittrouer, 1988).

Two of the results from the estimates of phrase familiarity—that is, median occurrence and minimum occurrence—are plotted in Figures 4b and 4c. These two estimates had the highest correlation with speechreading

difficulty: .87 and .92, respectively. The estimates drop systematically as a function of the list number—that is, sentences that are easy to speechread are characterized by phrases that are relatively common, and sentences that are difficult to speechread are characterized by phrases that are relatively less common. This result suggests that phrase familiarity plays a significant role in speechreading accuracy at the sentence level.

Experiment 4: Does Visual Information Limit the Value of Bimodal or Bilateral Stimulation?

It is well established that, in A test environments, low-frequency acoustic hearing in the ear contralateral to a CI (bimodal CI) provides information that significantly improves speech understanding relative to a single CI, especially in noise (e.g., Ching, van Wanrooy, & Dillon, 2007; Dorman et al., 2015). It is also well established that, in A test environments, a second implant (bilateral CIs) improves speech understanding in complex listening environments in which binaural cues can be exploited (e.g., Mosnier et al., 2009; Ricketts, Grantham, Ashmead, Haynes, & Labadie, 2006; Schön, Müller, & Helms, 2002).

In Experiment 2 we found that vision added 30–45 percentage points to performance for CI users listening to speech in noise. This outcome suggests the possibility that, in many common listening environments, AV performance will be near ceiling for speech understanding. For example, in Experiment 2, for the easy lists, AV performance in noise approached 90% correct. If, in common AV listening environments, performance is near ceiling, then bimodal hearing or bilateral CIs, known to have significant value in A listening environments, may be of little or no value for speech understanding.

In Experiment 4 we asked whether low-frequency acoustic hearing adds to the intelligibility provided by a single CI when V cues are available and whether a second CI benefits a single CI when V cues are available.

Method

Subjects

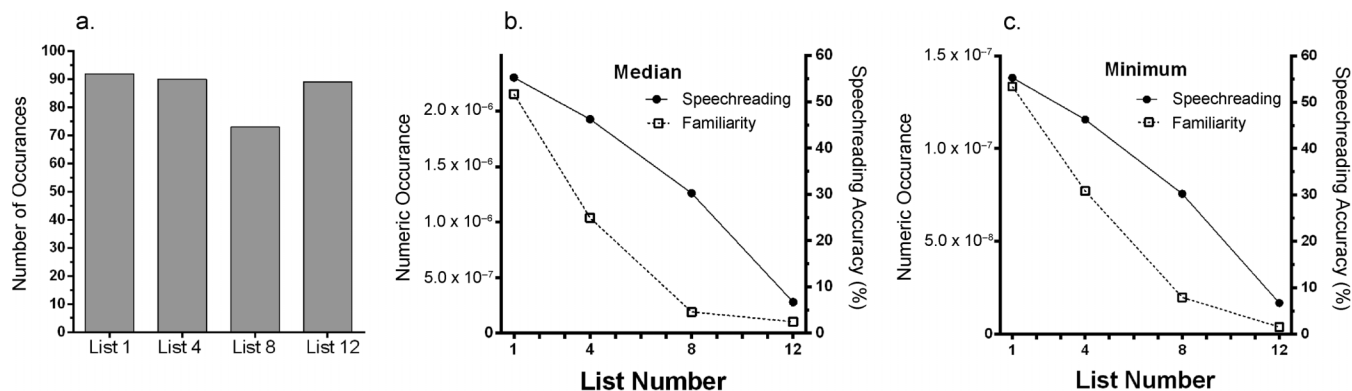
Seventeen bimodal CI users with postlingual deafness (six men, 11 women; age range = 34–83 years; mean age = 68 years) participated in the study. The mean unaided thresholds for the acoustically stimulated ear of these subjects at 125, 250, 500, and 750 Hz were, respectively, 47.2, 51.7, 71.1, and 80 dB HL. All patients wore hearing aids (HAs) set to match Verifit targets (Yehudai, Shpak, Most, & Luntz, 2013).

Four bilateral CI users with postlingual deafness (two men, two women; age range = 39–69 years; mean age = 54 years) participated in this study. All subjects had at least 2 years' experience with their CIs.

Material

The female-voice AzAV sentence lists were used as the test material.

Figure 4. (a) The summed number of occurrences of the consonants /m/, /b/, /p/, /w/, /s/, /ʃ/, /θ/, /ð/, /f/, and /v/ in sentence lists that vary in speechreading difficulty from easy (List 1) to difficult (List 12). (b) The estimated median, and (c) minimum occurrences of 3-grams (three-word sequences) in the Kopra lists (referenced to Google Ngrams) as a function of list difficulty are shown by the open squares and dotted line. Speechreading accuracy, reproduced from Figure 2, is shown by the filled circles and solid line.



Procedure

For the bimodal CI users, the test environment was the same as in the previous experiments—that is, both target and noise were presented from a single loudspeaker. In the CI-only and CI+V test conditions, the acoustically stimulated ear was plugged and muffled. As in Experiment 2b, a small set of sentences were presented in noise to estimate, for each listener, the speech-to-babble level necessary to drive performance to near 40% correct. This level, ranging from +10 to +3 dB, was then used with the test material.

The bilateral patients were seated in the R-space test environment (Compton-Conley, Neuman, Killion, & Levitt, 2004)—that is, seated in the middle of a circular array of speakers. Only three loudspeakers were used: Female-voice target signals from the AzBio sentences were presented from a speaker at 0°; one set of male-voice sentences (taken from the IEEE sentences; Rothausser et al., 1969) were concatenated, looped, and presented continuously from a speaker at +90°; and another set of sentences spoken by a different male speaker were delivered from a speaker at -90°. This test environment was chosen so that binaural cues relevant to spatial separation of target and maskers might be used.

For the bimodal CI users, CI-only, CI+HA, CI+V, and CI+V+HA sessions were tested in random order. For bilateral CI users, both CIs were first tested in isolation with CVC words and AzBio sentences to determine the poorer and better sides, using both scores combined. As was the case for the bimodal users, a small set of sentences were presented in noise to first estimate, for each listener, the babble level necessary to drive performance to near 40% correct. This level was then used with the test material. The better CI+V and bilateral CI+V conditions were then tested in random order. Two lists of the AzAV sentences were presented in each test condition. A short practice session preceded each test condition.

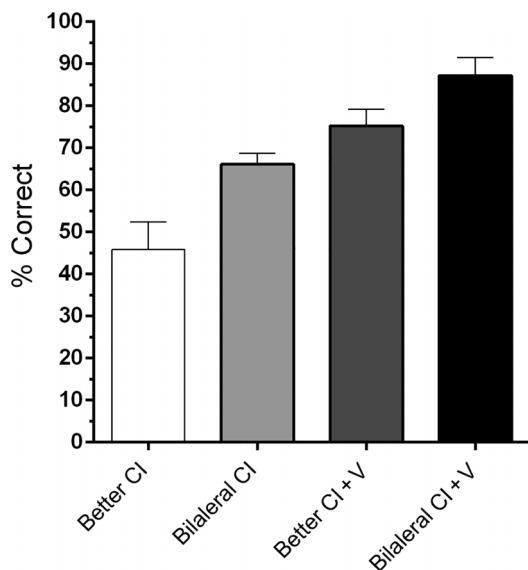
Results and Discussion

The results for the bilateral CI users are shown in Figure 5. The mean score in the better CI condition was 46% correct ($SEM = 7$); in the bilateral CI condition, 66% correct ($SEM = 3$); in the better CI+V condition, 75% correct ($SEM = 8$); and in the bilateral CI+V condition, 87% correct ($SEM = 9$). Three of the four subjects had higher scores in the better CI+V condition than in the bilateral CI condition, indicating that vision provided more information than the second CI in this listening environment. Nonetheless, three of the four subjects achieved higher scores in the bilateral CI+V condition than in the better CI+V condition, indicating that even when visual information was available, bilateral CIs were of value to speech understanding. The one subject who did not show better scores had minimal benefit from the second CI in the A conditions.

The results for the bimodal CI users are shown in Figures 6a–6c. Inspection of the raw data indicates that these subjects can be sorted into three subgroups on the basis of outcomes. As shown in Figure 6a, four subjects showed no gain in speech understanding when information from their HA was added to the information available in the CI+V condition—CI-only: 44% correct ($SEM = 5$); CI+HA: 69% correct ($SEM = 7$); CI+V: 85% correct ($SEM = 3$); CI+V+HA: 82% correct ($SEM = 5$). This outcome fits our supposition that information from vision could, in some environments, cause information from low-frequency acoustic hearing to be of little or no value for CI speech understanding (at least when scored in terms of percent correct).

In contrast to this outcome, as shown in Figure 6b, another subgroup ($n = 6$) showed a mean gain of 23 percentage points—CI-only: 34% correct ($SEM = 6$); CI+HA: 53% correct ($SEM = 7$); CI+V: 57% correct ($SEM = 8$); CI+V+HA: 80% correct ($SEM = 4$). Thus, low-frequency acoustic hearing can, in some conditions, be of benefit to CI users even when visual information is available.

Figure 5. Percent correct word recognition as a function of test condition for users of bilateral CIs. CI = cochlear implant; V = vision; error bars = ± 1 SEM.

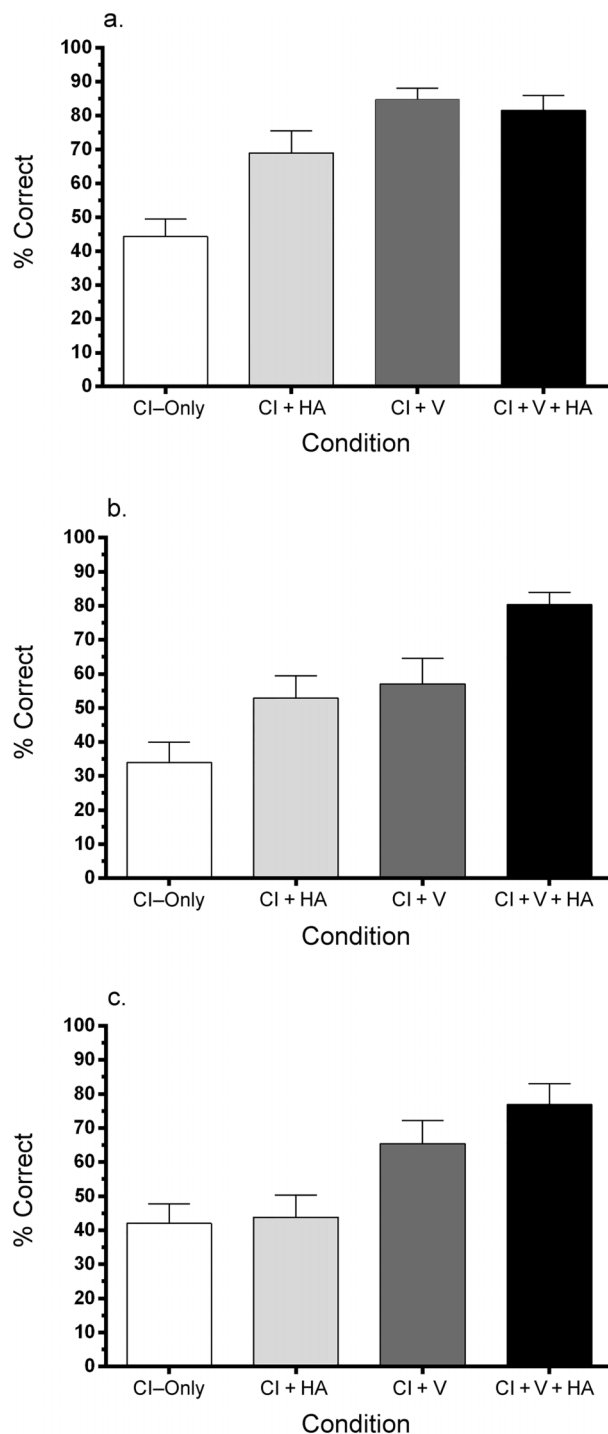


Inspection of Figure 5 indicates that two factors distinguished the two groups of subjects. One is the CI-only score—34% correct for the subjects who did benefit versus 44% correct for those who did not. The second factor is the gain provided by visual information. The subjects who did benefit gained 34 percentage points and had a mean CI+V score of 57% correct. Those who did not benefit gained 41 percentage points in performance from visual information and had a mean CI+V score of 85% correct. Thus, a combination of slightly better CI-only score and a slightly larger gain from visual information conspired to negate the usefulness of low-frequency acoustic information for speech understanding.

We note that there was room for improvement in the scores for the subjects who did not show benefit from adding low-frequency acoustic information—the mean score of 85% correct was not at ceiling. The absence of improvement, at this level of performance, indicates that the information from low-frequency hearing did not provide novel information to drive lexical access. In contrast, with a baseline of 57% correct, information from low-frequency acoustic hearing did provide novel information for lexical access.

As shown in Figure 6c, a third subgroup of subjects ($n = 7$) was characterized by no gain in speech understanding when the HA was added to the CI. This is in contrast to the two other subgroups, for whom the HA provided significant benefit when added to the CI. It is critical, however, that the HA did add to intelligibility when it was added to the information available from the CI+V. Thus, for this subgroup, information from vision enabled the information from low-frequency acoustic hearing to be of value. In the final experiment in this series, we explored

Figure 6. Percent correct word recognition in sentences as a function of test condition for CI users who had a low-frequency acoustic hearing aid (HA) in the contralateral ear. (a) Performance of four subjects who did not benefit from the HA when visual information was available. (b) Performance of six subjects who did benefit from the HA when visual information was available. (c) Performance of seven subjects who did not benefit from the HA when it was added to the CI but did when visual information was available. CI = cochlear implant; HA = low-frequency acoustic hearing aid; V = vision; error bars = ± 1 SEM.



the role played by vision for CI users, when other information is very poor, in providing information about lexical boundaries via use of the Metrical Segmentation Strategy (MSS; Cutler & Butterfield, 1992; Cutler & Norris, 1988; Liss, Spitzer, Caviness, Adler, & Edwards, 1998).

Experiment 5: Visual Information Promotes Syllabic Identification and Lexical Segmentation

Our findings from Experiment 3, linking phrase familiarity with visual benefit, are in line with the view that listener expectation is a driving mechanism in speech perception, particularly under challenging listening conditions (e.g., Mattys, White, & Melhorn, 2005). These expectations, importantly, operate at the word and phrase level.

In order to profitably use word- and phrase-sized units, listeners must be able to identify word boundaries within the connected speech stream. In English, the presence of a strong syllable—one emphasized in terms of duration and loudness and containing a full, unreduced vowel—most commonly signals a word onset (Cutler & Carter, 1987). Cutler and Carter have suggested that, when other information is reduced or unavailable, attention to the juxtaposition of strong and weak syllables could serve as a means for the identification of word boundaries. This strategy is referred to as the Metrical Segmental Strategy.

Recent literature suggests that vision can provide information about speech in a similar fashion to that provided by audition (e.g., Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Peelle & Sommers, 2015; Scarborough, Keating, Mattys, Cho, & Alwan, 2009; Swerts & Krahmer, 2008). For example, Chandrasekaran, Trubanova, Stillitano, Caplier, and Ghazanfar (2009) conducted a statistical analysis of the relationship between visual information and speech-envelope information and found robust correlations, among others, between the area of mouth opening (interlip difference) and the wideband power of the acoustic envelope. Thus, visual information can signal the presence of strong syllables and, in so doing, provide information about potential word onsets.

To explore this possibility for CI users, in Experiment 5 we recorded, in AV format, a set of specially constructed phrases that have been used in previous experiments to probe strategies for lexical segmentation (e.g., Liss et al., 1998; Liss, Spitzer, Caviness, & Adler, 2002). Unlike in the English language, where strong syllables predominate as word onsets, the distribution of strong and weak word onsets in this set of phrases is roughly balanced. By examining the type of lexical boundary error—insertion (I) or deletion (D)—and the location of the error—before a strong (S) or a weak (W) syllable—it is possible to determine whether or not listeners are attending to syllable strength to segment the acoustic stream into words. If listeners are using the MSS, their lexical-boundary errors will be predominantly erroneous insertions before strong syllables (IS) and erroneous deletions predominantly before weak syllables (DW). This analysis, along with a number of other

measures detailed in Data Analyses, allowed us to determine whether the addition of vision enhances reliance on the MSS, presumably secondary to the enhanced cues to syllable strength provided by vision.

Method

Subjects

Eight bilateral CI users with postlingual deafness (five women, three men; age range = 37–87 years; mean age = 68 years) participated. All subjects had used their implants for at least 1 year.

Material

A total of 160 three-to-five-word phrases consisting of six syllables with alternating stress, previously generated by Liss et al. (1998), were presented to each subject. The phrases were specially constructed to be of low interword probability, such that listeners would not be able to use semantic predictions to augment intelligibility. Half the phrases contained the S–W syllabic stress pattern characteristic of English (e.g., “balance clamp and body”) and the other half contained the less-common W–S stress pattern (e.g., “create her spot of art”). The phrases were semantically anomalous but syntactically plausible. They were recorded by a male speaker in AV format for this experiment.

Procedure

Testing was conducted in the single-loudspeaker AV test environment described for Experiment 1. The analysis of lexical-boundary errors requires that listeners make a large number of transcription errors. For that reason, for each listener the signal-to-babble ratio was adjusted in an A test condition (target signal at 60 dB SPL) to drive performance to 30%–40% correct. These values ranged from +15 to +3 dB. The subjects then completed short practice sessions using both A and AV presentation at that ratio.

During the experiment, each subject was presented with all 160 phrases. Phrases were divided into eight blocks of 20 phrases. Each block consisted of 10 phrases with the S–W syllabic stress pattern and 10 with W–S. Four blocks of 20 phrases were presented with auditory input only and four were presented with auditory and visual input. The order of presentation of A phrase blocks and AV phrase blocks was randomized. The subjects typed their responses on a keyboard for data entry.

Data Analyses

Three types of analyses were conducted: a measure of total words correct, an analysis of phonemic errors for consonants, and multiple measures of lexical segmentation. Response words that were analyzed for consonant errors were only those without lexical-boundary errors, syllable deletions, or syllable additions. Lexical segmentation measures included the total number of word-boundary errors, insertion of a word boundary before a strong (IS) or a weak (IW) syllable, deletion of a word boundary before a strong (DS) or a weak (DW) syllable, and whole-syllable

insertions or deletions that resulted in, respectively, more or less than six syllables.

Results and Discussion

As shown in Figure 7a, the mean word score for the CI-only condition was 37% correct ($SEM = 6$); for the CI+V condition the mean score was 54% correct ($SEM = 7$). The mean improvement in performance was 17 percentage points. As expected, the visual benefit was much smaller than we found in our other experiments (e.g., 44 percentage points in Experiment 2), because of the semantically anomalous stimuli. This is consistent with the results of Experiment 3, wherein visual benefit varies with familiarity of the word sequences in the test material. Nonetheless, the difference between the two scores differed significantly, $t(7) = -9.75$, $p < .001$, showing visual benefit for these phrases.

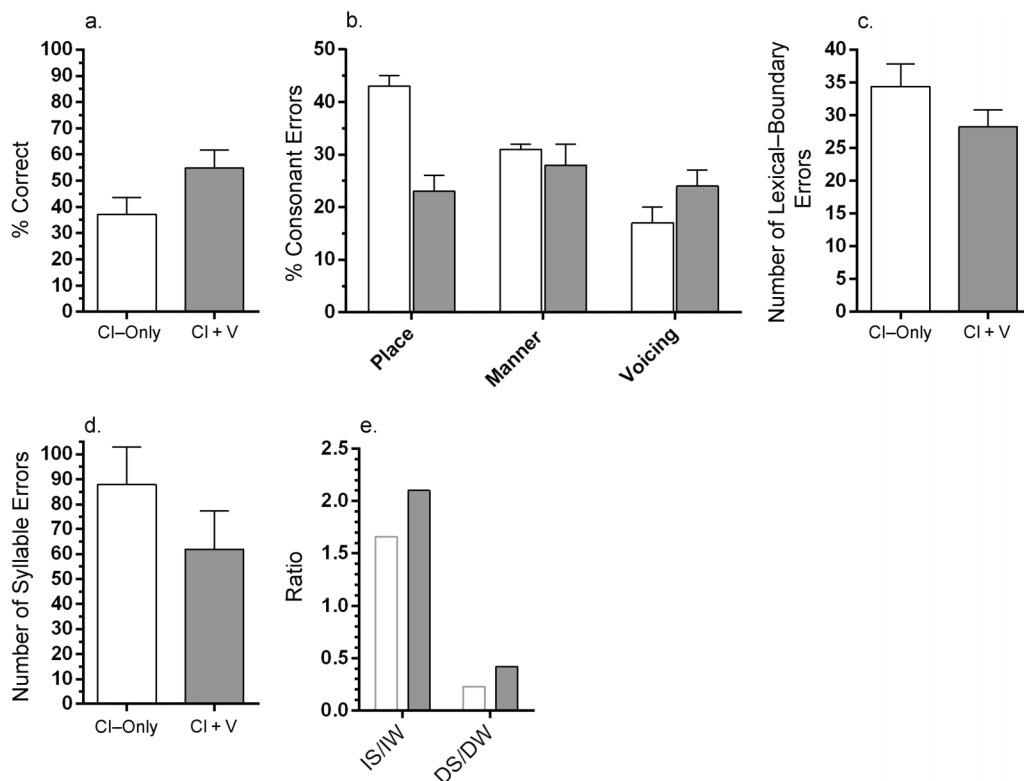
To assess the source of the visual benefit, we first examined the contribution of place of articulation within transcribed words that were in error but correctly lexically segmented (e.g., “taker” for the target “baker”). As shown in Figure 7b, adding visual information decreased the number of consonant-place error, $t(7) = 1.05$, $p < .01$; did not alter the number of manner error, $t(7) = 1.005$, $p > .01$;

and, curiously, increased the number of voicing errors, $t(7) = -3.2765$, $p < .01$. Thus, as expected, visual information significantly benefited only place of articulation.

The next set of analyses dealt with words that were incorrectly lexically segmented—that is, they violated a lexical boundary. As shown in Figure 7c, with the addition of visual information the mean number of lexical-boundary errors decreased from 34 ($SEM = 3$) to 28 ($SEM = 3$) across participants. The effect was marginally significant, $t(7) = 2.19$, $p = .064$, and probably can be attributed to the improvement in overall intelligibility with vision. That is, the more words correctly transcribed, the fewer opportunities for lexical-boundary errors to occur.

The most compelling explanation of our results is derived from the analyses of the lexical-boundary errors and is shown in Figure 7e. Recall that the MSS predicts that listeners will treat strong syllables as word onsets when they are unsure, and will therefore produce predominantly lexical-boundary insertions before strong syllables (IS). We calculate this as the IS/IW ratio, or the number of word-boundary insertions before a strong syllable divided by the number of word-boundary insertions before a weak syllable. If there is no use of the strategy, the ratio is close to 1.0, which mirrors chance, given that the corpus of

Figure 7. (a) Percent correct word recognition in CI and CI+V test conditions for phrases with low interword probabilities. (b) Percent total consonant errors in CI and CI+V test conditions for place, manner, and voicing. (c) Number of lexical-boundary errors in CI and CI+V test conditions. (d) Number of syllable-insertion and -deletion errors in CI and CI+V test conditions. (e) IS/IW and DS/DW ratios for CI-only and CI+V conditions. CI = cochlear implant; V = vision; IS = insertion before a strong syllable; IW = insertion before a weak syllable; DS = deletion before a strong syllable; DW = deletion before a weak syllable.



phrases is deliberately designed to provide roughly equal opportunities to generate IS and IW errors. The greater the positive deviation from 1.0, the more robust the treatment of strong syllables as word onsets.

For the CI-only condition, the IS/IW ratio was 1.66, suggesting some use of the MSS—an outcome consistent with a previous report from our laboratory (Spitzer, Liss, Spahr, Dorman, & Lansford, 2009). However, the addition of vision in the CI+V condition boosted the IS/IW ratio to 2.1. Thus, insertions occurred twice as often before strong syllables as before weak syllables, indicating increased use of the strategy with the presence of visual cues. A similar trend was found for the smaller number of deletion errors, in which lexical-boundary deletions before weak syllables outnumbered those before strong syllables, as predicted by the MSS. A nonparametric goodness-of-fit chi-square analysis comparing the four error type rates (IS, IW, DS, DW) between the two conditions (CI-only, CI+V) found a significant effect, $\chi^2(3, 285) = 8.83, p = .032$, indicating that the error-rate values from the two conditions were drawn from two different populations. This confirms that at least part of the intelligibility gain afforded by vision was due to an enhanced ability to identify syllable-strength cues, and therefore word onsets, for lexical segmentation.

A final analysis further confirms syllable-level benefits with the addition of vision. All target phrases were designed to be six syllables in length, but with number of words from three to five per phrase. The lexical-boundary error analysis already described focused on incorrect segmentations among the target six syllables. However, the CI users frequently produced transcription responses with either more or less than six syllables. We therefore tallied syllable additions and deletions as another measure of lexical-boundary identification. As shown in Figure 7d, erroneous syllable additions and deletions dropped substantially when visual information was available—CI-only: 88 ($SEM = 15$); CI+V: 62 ($SEM = 15$). These scores differed significantly, $t(7) = 2.9, p = .02$.

Our results taken together indicate that, as expected, visual information enhanced place-of-articulation cues for phoneme identification. However, visual information also facilitated lexical segmentation by improving the recognition of the number of syllables produced and the relative strength of these syllables. As Auer (2009) has noted, successful lexical segmentation is a prerequisite for the use of lexical constraints in both auditory and visual speech recognition. Our contribution to this literature is a demonstration that visual information can aid in lexical segmentation.

General Discussion

The information derived from the five experiments reported here (a) provides answers to questions about levels of performance to be expected in common listening environments for CI users and (b) points to mechanisms underlying the increase in speech understanding when auditory and visual information is available to CI users.

Levels of Performance

In Figures 8a–8c we summarize data, collected in Experiments 2a and 4, on the gain in speech understanding using AV stimulation. In the A conditions, subjects could have a single CI, bilateral CIs, or a bimodal fitting. As shown in Figure 8a, the mean improvement was 30 percentage points. As shown in Figure 8b, few subjects had a gain of less than 10 percentage points or greater than 40. Figure 8c indicates that gain from V information is constrained by the A score: The largest gain occurs when A scores are in an intermediate range (see also Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). This is reasonable—when A scores are very low, phonetic information is severely degraded and additional information about syllable strength and lexical boundaries is of small value for constraining lexical choices. When A scores are very high, on the other hand, there are multiple sources of information about phonetic and lexical identity, and then information from vision provides minimal additional information to constrain and improve lexical choice. That said, with the exception of one outlier, vision added at least 10 percentage points to speech understanding over the range of A scores from 10%–75% correct.

In our AV test conditions, subjects needed to integrate information from multiple sources. Many models have been proposed to account for the interaction of information sources for speech understanding. For recent models relevant to the interaction of electric and acoustic stimulation, see Micheyl and Oxenham (2012) and the references therein.

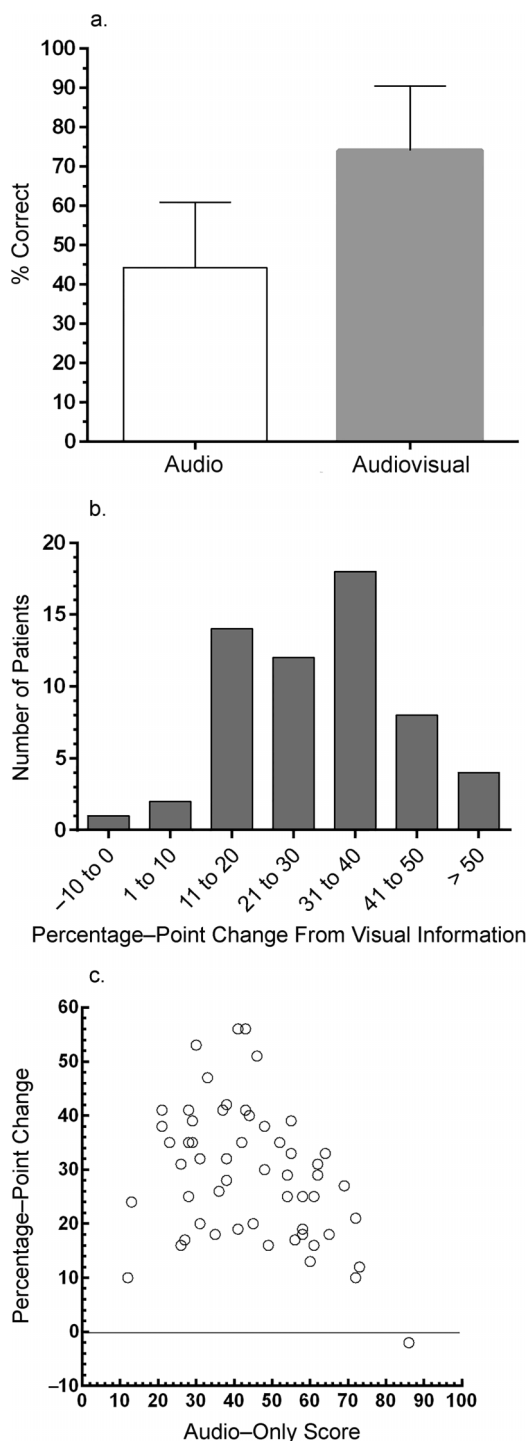
The Value of a Bimodal CI and a Second CI When Visual Information Is Present

As we noted earlier, many studies have documented the value in A test conditions of bilateral CIs, and of low-frequency acoustic hearing, in terms of improving performance relative to a single CI. In Experiment 4 we asked whether testing in A environments has misled us about the value of these two interventions in the most common listening environment for CI users—that is, when both A and V information is available.

The mean CI-only score for a large number of CI users with postlingual deafness tested in quiet with the relatively easy HINT sentences (Nilsson, Soli & Sullivan, 1994) is approximately 85% correct (Gifford et al., 2008). For a large number of CI users with postlingual deafness tested with the more difficult AzBio sentences, the mean score is approximately 70% correct (Wilson, Dorman, Gifford, & McAlpine, 2016). If visual information is available in a quiet environment, then neither bilateral nor bimodal CIs are likely to be of additional value to speech understanding. We note that the addition of a low-frequency acoustic signal or a second CI may affect ease of listening and/or sound quality in an AV environment, but our data do not speak to those issues.

To gauge the value of low-frequency acoustic hearing or a second CI when visual information is available,

Figure 8. Overall results from subjects tested in A and AV conditions in Experiments 2b and 4. (a) Mean percent correct in A and AV conditions. Error bars = ± 1 SD. (b) Percentage-point change from visual information. (c) Percentage-point change in AV condition as a function of A score.



we should look to environments in which noise is present. For bilateral CI users, our findings were relatively straightforward: A second CI was of benefit in environments in which binaural cues could be exploited when visual information was present—even when the best CI+V score was above 80% correct. For CI users with low-frequency acoustic hearing, the outcomes were more nuanced. Subjects with relatively good CI performance in noise and relatively large benefit from vision did not benefit from adding low-frequency acoustic information. Subjects with relatively poor CI performance and relatively little gain from vision did benefit. The performance of a third subgroup was the most interesting: Low-frequency acoustic hearing did not improve CI-only performance but did improve CI+V performance. Thus, visual information facilitated or potentiated the use of the information provided by low-frequency acoustic hearing.

The Role of Visual Information in Improving CI Performance

Peelle and Sommers (2015) reviewed the current understanding of AV speech perception as an integrated multisensory phenomenon within a framework of “temporally-focused lexical competition” (p. 169). They extracted evidence for this view from the body of literature which demonstrates that both predictive and constraining mechanisms drive AV benefits (e.g., Chandrasekaran et al., 2009; Girard & Poeppel, 2012; Peelle & Davis, 2012; Tye-Murray, Spehar, Myerson, Sommers, & Hale, 2011). To summarize their primary-literature review: Visual information from tracking mouth movement during speech coincides with the amplitude envelope of the speech signal. Neural oscillations entrain to these rhythms, roughly at the level of the syllable, which supports higher neural sensitivity to acoustic information occupying these intervals. This temporal focus serves to reduce cognitive load and, in turn, improve speech-processing efficiency. Further, the visual information enhances phoneme recognition, thereby constraining lexical competitors and facilitating lexical access. Through both earlier- and later-stage multisensory integration, the system leverages rhythmic predictability and lexical constraints to locate and identify words in connected speech.

Our present data on CI lexical segmentation fit well within this interpretive framework. The AV condition ostensibly allowed for enhanced tracking of amplitude-envelope fluctuations through tracking mouth movements and other synchronized movements of the head, face, or body. This facilitated the tasks of identifying syllables in the acoustic stream and recognizing their relative juxtaposed strengths. The data on lexical-boundary error from Experiment 5 suggest that CI users used this enhanced syllable resolution (perhaps due to vision enhancing the sharpness of the acoustic fluctuations across time; Doelling, Arnal, Ghitza, & Poeppel, 2014), along with lexical constraints imposed by visible phoneme information, to better identify word onsets and, in so doing, to improve lexical access.

Notes on Limits of the Data Set

Our initial project was a survey in which we asked CI users about the environments in which they listened to speech. In the absence of instrumental data—for example, a head-mounted video camera and microphone—we cannot confirm the accuracy of the answers. It would be of some interest if subjects' introspections about listening environments were substantially in error.

One of the motivations of this work was to provide an answer to the question, asked by CI users before surgery, "How well will I be able to understand speech?" A comprehensive answer to the question would require testing A and AV speech understanding in multiple test environments, each representative of a common listening environment. We have not done that, and we make no claim that our results represent results from all, or even most, listening environments for CI users.

We used only one female speaker in our experiments. Multiple speakers of different genders would, of course, be more representative. That said, listeners attend most commonly to one speaker at a time in complex listening environments. A test environment in which there is a new speaker from trial to trial (or sentence to sentence) does not, in our experience, represent a common, real-world listening environment.

In Experiment 4 only four bilateral CI users were tested. This is a very small sample, and the results should be interpreted as preliminary. Nonetheless, our results, indicating the benefit of two CIs versus one when visual information is present, are consistent with data in a recent article by van Hoesel (2015).

Acknowledgments

This article reports the results from the PhD dissertation of Shuai Wang and the master's-thesis work of Cimarron Ludwig. Experiment 3 was entirely the work of Shuai Wang and Visar Berisha. These projects were supported by National Institute on Deafness and Other Communication Disorders Grants R01 DC 010821 and R01 DC DC010494, awarded to Michael F. Dorman, and R01 DC006859, awarded to Julie Liss.

References

- Altieri, N. A., Pisoni, D. B., & Townsend, J. T. (2011). Some normative data on lip-reading skills (L). *The Journal of the Acoustical Society of America*, *130*, 1–4.
- Auer, E. T. (2002). The influence of the lexicon on speech read word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin & Review*, *9*, 341–347.
- Auer, E. T. (2009). Spoken word recognition by eye. *Scandinavian Journal of Psychology*, *50*, 419–425.
- Bergeson, T. R., Pisoni, D. B., Reese, L., & Kirk, K. I. (2003, February). *Audiovisual speech perception in adult cochlear implant users: Effects of sudden vs. progressive hearing loss*. Poster presented at the Annual MidWinter Research Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. Sebastopol, CA: O'Reilly Media.
- Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*, *84*, 101–114.
- Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 1001–1010.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLOS Computational Biology*, *5*(7), e1000436.
- Ching, T. Y. C., van Wanrooy, E., & Dillon, H. (2007). Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends in Amplification*, *11*, 161–192.
- Clinkard, D., Shipp, D., Friesen, L. M., Stewart, S., Ostroff, J., Chen, J. M., . . . Lin, V. Y. W. (2011). Telephone use and the factors influencing it among cochlear implant patients. *Cochlear Implants International*, *12*, 140–146.
- Compton-Conley, C. L., Neuman, A. C., Killion, M. C., & Levitt, H. (2004). Performance of directional microphones for hearing aids: Real-world versus simulation. *Journal of the American Academy of Audiology*, *15*, 440–455.
- Cook, S. (2014, June). *Evaluating speech perception ability using new audio-visual test material*. Poster presented at the 13th International Conference on Cochlear Implants and Other Implantable Auditory Technologies, Munich, Germany.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*, 133–142.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.
- Desai, S., Stickney, G., & Zeng, F.-G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America*, *123*, 428–440.
- Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768.
- Dorman, M. F., Cook, S., Spahr, A., Zhang, T., Loiselle, L., Schramm, D., . . . Gifford, R. (2015). Factors constraining the benefit to speech understanding of combining information from low-frequency hearing and a cochlear implant. *Hearing Research*, *322*, 107–111.
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*, 502–504.
- Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology & Neurotology*, *13*, 193–205.
- Girard, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517.
- Gray, R. F., Quinn, S. J., Court, I., Vanat, Z., & Baguley, D. M. (1995). Patient performance over eighteen months with the Ineraid intracochlear implant. *Annals of Otolaryngology, Rhinology & Laryngology Supplement*, *166*, 275–277.
- Humes, L. E. (1991). Understanding the speech-understanding problems of the hearing impaired. *Journal of the American Academy of Audiology*, *2*, 59–69.

- Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research, 46*, 390–404.
- Kirk, K. I., Prusick, L., French, B., Gotch, C., Eisenberg, L. S., & Young, N. (2012). Assessing spoken word recognition in children who are deaf or hard of hearing: A translational approach. *Journal of the American Academy of Audiology, 23*, 464–475.
- Kopra, L. L., Kopra, M. A., Abrahamson, J. E., & Dunlop, R. J. (1986). Development of sentences graded in difficulty for lipreading practice. *Journal of the Academy of Rehabilitative Audiology, 19*, 71–86.
- Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112*, 3022–3030.
- Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*, 2457–2466.
- Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*, 131–141.
- 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.
- Massa, S. T., & Ruckenstein, M. J. (2014). Comparing the performance plateau in adult cochlear implant patients using HINT and AzBio. *Otology & Neurotology, 35*, 598–604.
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477–500.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, . . . Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*, 176–182.
- Micheyl, C., & Oxenham, A. J. (2012). Comparing models of the combined-stimulation advantage for speech recognition. *The Journal of the Acoustical Society of America, 131*, 3970–3980.
- Mosnier, I., Sterkers, O., Bebear, J.-P., Godey, B., Robier, A., Deguine, O., . . . Ferrary, E. (2009). Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients. *Audiology & Neurotology, 14*, 106–114.
- Most, T., Rothem, H., & Luntz, M. (2009). Auditory, visual, and auditory-visual speech perception by individuals with cochlear implants versus individuals with hearing aids. *American Annals of the Deaf, 154*, 284–292.
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science, 15*, 133–137.
- Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*, 1085–1099.
- Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*, 320.
- Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex, 68*, 169–181.
- Ricketts, T. A., Grantham, D. W., Ashmead, D. H., Haynes, D. S., & Labadie, R. F. (2006). Speech recognition for unilateral and bilateral cochlear implant modes in the presence of uncorrelated noise sources. *Ear and Hearing, 27*, 763–773.
- Robbins, A. M., Renshaw, J. J., & Osberger, M. J. (1995). *Common Phrases Test*. Indianapolis, IN: Indiana University School of Medicine.
- Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 51–78). Malden, MA: Blackwell.
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science, 17*, 405–409.
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*, 1147–1153.
- Rothauer, E., Chapman, W., Guttman, N., Hecker, M., Nordby, K., Silbiger, H., Urbanek, G., & Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17*, 225–246.
- Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Research, 1188*, 87–99.
- Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech, 52*, 135–175.
- Schön, F., Müller, J., & Helms, J. (2002). Speech reception thresholds obtained in a symmetrical four-loudspeaker arrangement from bilateral users of MED-EL cochlear implants. *Otology & Neurotology, 23*, 710–714.
- Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 18748–18750.
- Spahr, A., Dorman, M. F., Litvak, L. M., Cook, S., Loiseau, L. M., DeJong, M. D., . . . Gifford, R. H. (2014). Development and validation of the Pediatric AzBio sentence lists. *Ear and Hearing, 35*, 418–422.
- Spahr, A., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., . . . Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and Hearing, 33*, 112–117.
- Spitzer, S., Liss, J., Spahr, T., Dorman, M., & Lansford, K. (2009). The use of fundamental frequency for lexical segmentation in listeners with cochlear implants. *The Journal of the Acoustical Society of America, 125*, EL236–EL241.
- Strelnikov, K., Rouger, J., Barone, P., & Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian Journal of Psychology, 50*, 437–444.
- Summerfield, A. Q. (1985). Speech-processing alternatives for electrical auditory stimulation. In R. A. Schindler & M. M. Merzenich, *Cochlear implants* (pp. 195–222). New York, NY: Raven Press.
- Summerfield, A. Q. (1987). Some preliminaries to a theory of audiovisual speech processing. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 58–82). Hillsdale, NJ: Erlbaum.
- Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics, 36*, 219–238.
- Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., & Rose, N. S. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology, 47*(Suppl. 2), S31–S37.

-
- Tye-Murray, N., Spehar, B., Myerson, J., Sommers, M. S., & Hale, S.** (2011). Cross-modal enhancement of speech detection in young and older adults: Does signal content matter? *Ear and Hearing, 32*, 650–655.
- Tyler, R. S., Preece, J., & Lowder, M.** (1983). *The Iowa Cochlear Implant Tests*. Iowa City, IA: University of Iowa Department of Otolaryngology, Head and Neck Surgery.
- van Dijk, J. E., van Olphen, A. F., Langereis, M. C., Mens, L. H. M., Brokx, J. P. L., & Smoorenburg, G. F.** (1999). Predictors of cochlear implant performance. *International Journal of Audiology, 38*, 109–116.
- van Hoesel, R.** (2015). Audio-visual speech intelligibility benefits with bilateral cochlear implants when talker location varies. *Journal of the Association for Research in Otolaryngology, 16*, 309–315.
- Wanna, G. B., Noble, J. H., Carlson, M. L., Gifford, R. H., Dietrich, M. S., Haynes, D. S., . . . Labadie, R. F.** (2014). Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes. *The Laryngoscope, 124*(S6), S1–S7.
- Wikman, A.** (2006). Reliability, validity and true values in surveys. *Social Indicators Research, 78*, 85–110.
- Wilson, R., Dorman, M., Gifford, R., & McAlpine, D.** (2016). Cochlear implant design considerations. In N. Young & K. I. Kirk (Eds.), *Pediatric cochlear implantation: Learning and the brain* (pp. 3–23). New York, NY: Springer.
- Yehudai, N., Shpak, T., Most, T., & Luntz, M.** (2013). Functional status of hearing aids in bilateral-bimodal users. *Otology & Neurotology, 34*, 675–681.