Published in final edited form as:
J Acoust Soc Am. 2008 January ; 123(1): 428–440. doi:10.1121/1.2816573.

Auditory-visual speech perception in normal-hearing and cochlear-implant listeners [a)]

Sheetal Desai, **Ginger Stickney**, and **Fan-Gang Zeng**[b)]
Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences and Otolaryngology - Head and Neck Surgery, 364 Medical Surgery II, University of California, Irvine, California 92697-1275

Abstract

The present study evaluated auditory-visual speech perception in cochlear-implant users as well as normal-hearing and simulated-implant controls to delineate relative contributions between sensory experience and cues. Auditory-only, visual-only, or auditory-visual speech perception was examined in the context of categorical perception, in which an animated face mouthing /ba/, /da/, or /ga/ was paired with synthesized phonemes from an 11-token auditory continuum. A 3-alternative, forced-choice, method was used to yield percent identification scores. Normal-hearing listeners showed sharp phoneme boundaries and strong reliance on the auditory cue, whereas actual and simulated implant listeners showed much weaker categorical perception but stronger dependence on the visual cue. The implant users were able to integrate both congruent and incongruent acoustic and optical cues to derive relatively weak but significant auditory-visual integration. This auditory-visual integration was correlated with the duration of the implant experience but not the duration of deafness. Compared with the actual implant performance, acoustic simulations of the cochlear implant could predict the auditory-only performance but not the auditory-visual integration. These results suggest that both altered sensory experience and improvised acoustic cues contribute to the auditory-visual speech perception in cochlear-implant users.

I. INTRODUCTION

Multi-sensory integration provides a natural and important means for communication. The benefit of integrating auditory and visual (AV) cues in speech perception has been well documented, particularly in difficult listening situations and for hearing-impaired listeners (Sumby and Pollack, 1954; Erber, 1972; Binnie *et al*., 1974; Dodd, 1977; Summerfield, 1979; Easton and Basala, 1982; Walden *et al*., 1993; Grant *et al*., 1998; Massaro, 1998). The benefit derived from speechreading can be substantial, allowing unintelligible speech to become comprehensive, or even exceeding the benefit derived from the use of assistive listening devices, counseling, or training (Sumby and Pollack, 1954; Walden *et al*., 1981; Montgomery *et al*., 1984; Grant and Braida, 1991; Grant and Walden, 1996). Two fundamental questions arise naturally concerning this AV integration in speech recognition: (1) what acoustic and optical cues are integrated, and (2) how and where are they integrated in the brain (Rosen *et al*., 1981; Braida, 1991; Massaro and Cohen, 2000; Bernstein *et al*., 2002; De Gelder and Bertelson, 2003)?

[a)] Portions of this work were presented at the Conference on Implantable Auditory Prostheses, Asilomar Conference Center, Pacific Grove, California, 2003

[b)] Send Correspondence to: Fan-Gang Zeng University of California, Irvine 364 Med Surge II Irvine, CA 92697 Phone: (949) 824-1539 FAX: (949) 824-5907 E-mail: fzeng@uci.edu.

Acoustic and optical cues can be complementary to enhance speech perception. On the one hand, some speech sounds can be more easily distinguished in the visual modality than in the auditory modality, e.g., the bilabial /ba/ versus the alveolar /da/ stop consonant (Binnie *et al*., 1974; Dodd, 1977; Summerfield, 1979; Walden *et al*., 1990). On the other hand, other speech sounds, called visemes, such as /b/, /p/, and /m/, rely on the acoustic cues to differentiate from each other because they are visually indistinguishable (Fisher, 1968; Binnie *et al*., 1974). While normal-hearing listeners have little trouble doing so, hearing-impaired listeners, including cochlear-implant listeners, often have great difficulty differentiating these phonemes due to reduced auditory temporal and spectral resolution (Miller and Nicely, 1955; Turner *et al*., 1997; Munson *et al*., 2003).

Because the optical cues provided by a speaker's facial and lip movements are not affected by the presence of noise, they are particularly useful at relatively low signal-to-noise ratios (SNRs). Sumby and Pollack (1954) demonstrated robust and relatively constant visual contribution to AV speech perception at SNRs over a 30-dB range. A recent study, however, showed the maximal integration efficiency at an intermediate SNR of about -12 dB (Ross *et al*., 2007).

There is also evidence that different acoustic cues contribute differently to the amount of auditory and visual integration. When presented acoustically, voice pitch is virtually unintelligible but can significantly improve speechreading. The reason for this improvement is that the voice pitch cue provides important segmental and suprasegmental information that is usually invisible (Rosen *et al*., 1981; Grant, 1987). Similarly, the temporal waveform envelope cue improves speechreading as long as the periodicity information is included (>50-500 Hz, see Rosen, 1992). If only the envelope information (<25 Hz) is included, then this temporal envelope cue produces little or no effect on speechreading (Grant *et al*., 1991; 1994). Overall, these studies suggest that it is not necessary to provide accurate information in both modalities, rather complementary acoustic and optical cues are sufficient to support high-level AV speech perception (e.g., Van Tasell *et al*., 1987).

To understand how and where these acoustic and optical cues are integrated in the brain, researchers have used incongruent cues from auditory and visual modalities (e.g., Calvert and Campbell, 2003; De Gelder and Bertelson, 2003; van Wassenhove *et al*., 2007). A compelling example showing interactions between incongruent auditory and visual cues in speech perception is the McGurk effect (McGurk and MacDonald, 1976). The McGurk effect is evoked by dubbing the audio recording of one sound (e.g., /ba/) onto the visual recording of a different sound (e.g., /ga/), obligating many listeners to report hearing an illusive sound (e.g., /da/ in this case). The McGurk effect has been extended to sentences, different languages, children, hearing-impaired listeners, and special patient populations (Green *et al*., 1991; Sams *et al*., 1998; Cienkowski and Carney, 2002; Burnham and Dodd, 2004). It is believed that AV speech integration occurs in a relatively early, pre-lexical integration stage (e.g., Calvert *et al*., 1997; Reale *et al*., 2007).

Recently, there has been an intensified interest in using the cochlear implant to study AV speech perception and integration. There are at least two reasons for this intensified interest. First, because the present implant extracts and delivers only the temporal envelope cue, lacking access to fine structure including the low-frequency voice pitch that is typically accessible to a hearing aid user, the implant users are particularly susceptible to noise (e.g., Stickney *et al*., 2004; Kong *et al*., 2005; Zeng *et al*., 2005). The optical cue, when available, is essentially unaffected by noise. Therefore, the implant users rely more than normal listeners on the visual cue, forcing them to become not only better speechreaders but also better multi-sensory integrators (Goh *et al*., 2001; Clark, 2003; Schorr *et al*., 2005; Rouger *et al*., 2007). Indeed, some cochlear-implant users can integrate AV cues to increase the functional SNR in noise

(Lachs *et al*., 2001; Bergeson *et al*., 2005; Hay-McCutcheon *et al*., 2005; Moody-Antonio *et al*., 2005).

Second, the dramatic auditory experience and intervention with the cochlear implant provide a unique tool to study brain plasticity in multiple ways. For example, Schorr et al. (2005) demonstrated a critical period in developing AV integration, with the critical age of implantation being at about 2.5 years old. On the other hand, brain imaging studies have shown a profound cortical reorganization in cochlear implant users, with good users being able to recruit a larger cortical area, even the visual cortex, than poor users to perform an auditory task (Giraud *et al*., 2001b; Lee *et al*., 2001; Doucet *et al*., 2006).

At present it remains unclear how much the cochlear-implant users can integrate auditory and visual information and whether this integration is related to stimulus and subject variables. The primary goal of the present study was to address the following two questions: (1) Do post-linguistically deafened persons fitted with a cochlear implant really integrate auditory and visual information? (2) How will altered stimuli and sensory experience affect AV integration? We first quantified the degree of AV integration by measuring performance in normal-hearing listeners, actual implant listeners, and simulated-implant listeners. We then delineated the relative contributions of stimulus and subject variables to AV integration by relating the degree of the AV integration to the duration of deafness and the duration of the implant experience.

II. METHODS

A. Subjects

1. Normal-hearing listeners—A total of 14 young, normal-hearing listeners participated in this study. All subjects were native English speakers with reported normal hearing. These young subjects ranged in age from 18 to 36 years. Subjects reported normal or corrected-to-normal vision.

Because of the large age difference between normal-hearing and cochlear-implant listeners as well as the known cognitive differences in sensory and cross-modality processing (Walden *et al*., 1993; Gordon-Salant and Fitzgibbons, 1997; Humes, 2002; Hay-McCutcheon *et al*., 2005), three elderly, nearly normal-hearing listeners (average age = 77) were recruited to evaluate whether age is a significant factor in the present study. Pure-tone averages (across 500, 1000 and 2000 Hz) of two subjects were below 15 dB HL, and the third subject had a pure-tone average of 22 dB HL. Pure-tone averages for all three subjects were taken from the right ear. These elderly subjects reported normal or corrected-to-normal vision. Because of time limitation, they only participated in the experiment with the original unprocessed stimuli.

2. Cochlear-Implant Listeners—A total of 8 post-lingually deafened, adult cochlear-implant listeners were evaluated in this experiment. These subjects were recruited locally from the Southern California area. They had a mean age of 66 years old, duration of deafness of 18 years, and >1 year of experience with their device. Table 1 shows additional information on the individual cochlear-implant listeners evaluated in this study, including consonant and vowel identification scores. Significant correlation was observed between duration of deafness and consonant (r=-0.96) and vowel (r=-0.85) recognition, as well as between implant experience and consonant (r=-0.70) and vowel (r=-0.91) recognition. All cochlear-implant listeners were native English speakers and were post-lingually deafened. All cochlear-implant listeners reported normal or corrected-to-normal vision[1].

[1] One of the cochlear-implant listeners tested reported that she had corrective cataract surgery and that the vision in the right eye was still somewhat limited. However, she also reported that glasses are sufficient for viewing the television and objects at close range. She was also able to drive during daylight hours.

B. Stimuli

1. Unprocessed Stimuli—The auditory stimuli were created using a web-based Klatt Synthesizer (1980), developed by Bunnell and colleagues at the Speech Research Laboratory, A.I. duPont Hospital for Children (1996). Eleven consonant-vowel (CV) tokens were synthesized to represent an auditory continuum along /ba/, /da/ and /ga/. The continuum was created by varying the starting frequency of the F2 or F3 formants in 200 Hz steps while keeping the other formant frequencies constant (see Table 2). The formant frequencies for the /a/ sound paired with each consonant were also kept constant. Reference tokens for /ba/, /da/ and /ga/ are highlighted in gray. The formant values for these reference tokens were adopted from Turner and Robb (1987).

To test the effects of formant-transition-duration (i.e. from the onset of the consonant sound to the onset of the steady state /a/ sound), two continuums were created: one with a formant-transition-duration of 20-ms and a second with a formant-transition-duration of 40-ms. The fundamental frequency for the consonant sounds started at 150 Hz and changed after 20- or 40-ms (depending on the type of stimulus) and then decreased to 100 Hz for the vowel sound over the remaining duration.

The total duration of each CV stimulus token was kept constant at 305-ms for both 20- and 40-ms stimuli (i.e. 20- or 40-ms was allotted to the respective consonant sounds and the remaining portion of the 305-ms was designated as the vowel sound). The auditory stimuli were calibrated using the Bruel & Kjaer sound level meter (Model No. 2260). A calibration tone, created by using a 1000 Hz sinusoid matched to the same rms level as the synthesized speech sounds, was used to adjust the level of the auditory stimuli to a 70 dB SPL presentation level.

Normal-hearing and cochlear-implant listeners were seated in a sound-attenuated booth during the experiments. Normal-hearing listeners listened to auditory stimuli monaurally through the right ear with Sennheiser HDA 200 headphones. Seven of the cochlear-implant listeners were presented with stimuli through a direct audio input connection to their speech processor and one cochlear-implant listener was presented with stimuli through a speaker because her ear-level device did not allow a direct audio connection.

Visual stimuli from an animated face (“Baldi”), which corresponded to the /ba/, /da/ and /ga/ sounds, were created using the Center for Spoken Language Understanding (CSLU) Speech Toolkit (Barnard *et al*., 2000; Massaro *et al*., 2000). The animated face was temporally aligned with each auditory stimulus token to represent the initial consonant position, the transition (20 or 40 ms), and the final vowel position. The computer monitor window displaying the animated face was modified so that the lips were 2" in width and 1" in height.

The synthetic sound and the animated face, instead of a natural sound and a human face, were used for the following two reasons. First, they can rule out possible confounding factor of idiosyncratic acoustic and optical cues in a small set of stimuli as used in the present study. Second, The “Baldi” program allowed accurate temporal alignment between the congruent and incongruent acoustic and optical cues. An apparent weakness of using these synthetic stimuli was their relatively weaker signal strength compared with natural stimuli, because the synthetic stimuli were limited to the number of variables in the models, resulting in lower accuracy and poorer resolution than the natural stimuli (Massaro *et al*., 2000). This weak signal strength could have contributed to relatively low-level AV integration (e.g., the McGurk effect) found in the present study.

2. Four- and 8-channel processed stimuli—The cochlear-implant simulation was generated using an algorithm developed by Shannon et al. (1995). The original stimuli from the 11-token continuum were band-pass filtered (6th-order elliptical IIR filters) by either 4- or

8 bands using the Greenwood map (1990). The envelope from each band was extracted through full-wave rectification followed by a 500-Hz low-pass filter (1st-order Bessel IIR filter). The envelope was then used to modulate a sinusoidal carrier set to the center frequency of the narrowband. The outputs of the narrowband signals were combined to create the 4- or 8-channel cochlear-implant simulations.

C. Procedures

We adopted a classical categorical perception paradigm similar to that used by Walden et al. (1990) in hearing-impaired listeners and by Clark and colleagues (2003) in pre-lingually deafened pediatric cochlear-implant users. A 3-alternative-forced-choice procedure was used in three experimental conditions: auditory-alone (A), visual-alone (V), and AV (AV). In the AV condition, an animated face mouthing /ba/, /da/, or /ga/ was paired with each speech sound from the auditory continuum, creating both congruent and incongruent AV combinations. The V condition evaluated the subjects' ability to lipread the mouthed phonemes. Each condition was tested with both 20- and 40-ms formant-transition-durations. Subjects were given both verbal and written instructions to click buttons labeled /ba/, /da/, and /ga/ on a computer-based interface that corresponded with what they thought they perceived. The 3-alternative forced-choice protocol allowed quantification of perceptual boundaries but could be a liability if the subject perceived a sound that was different from the three choices, e.g., a /bg/ response to an auditory /g/ and visual /b/ stimuli (McGurk and MacDonald, 1976).

Prior to the test session, practice sessions were given with feedback for the A, V, and AV conditions. For the A and AV conditions, the reference tokens were presented a total of 20 times, resulting in 60 presentations for one complete practice session. For the V practice sessions, each reference token was mouthed a total of 10 times, resulting in 30 presentations for one complete practice session. Normal-hearing listeners were required to achieve 80% correct identification of all three unprocessed CV pairs in the A practice session (at a 40-ms formant-transition-duration) to continue with the experiment. All of the normal-hearing listeners met this criterion.

Following the practice sessions, the test sessions were given without feedback. The test order of these conditions was randomized for each subject. For the A condition, each of the 11 continuum tokens was presented randomly a total of 20 times, resulting in 220 presentations for one complete test session. For the AV condition, each of the 11 continuum tokens was paired with a /ba/, /da/ and /ga/ face and were presented randomly a total of 10 times, resulting in 330 presentations. For the V condition, each /ba/,/da/ and /ga/ face was presented randomly a total of 20 times, resulting in 60 presentations for one complete test session. The scores were calculated as the number of tokens correctly identified (in % Correct) or as a distribution in response to each of the three tokens (in % Identification).

D. Data analysis

The phonemic boundaries were estimated with a 4-parameter sigmoidal function fit to each function[2]:

[2] Occasionally there were functions that did not optimally fit the parameters of the sigmoid 4-parameter equation. In these cases, alternative methods were used to find the x-intercepts: (1) if a 5-parameter sigmoidal function was able to fit a boundary instead of a 4-parameter sigmoidal function, then the values for a, b, x_0, and y_0 were taken from calculations made by the graphing software (SigmaPlot); (2) if two boundary lines were linear functions, the x-intercept was determined from the equation of the intersecting lines. When none of these rules applied, for instance when one function was linear and the second sigmoidal, the point that most closely estimated the x-intercept was taken upon inspection of the graph. These exceptions only occurred in 9 of 672 cases.

$$f(x) = y_0 + \frac{a}{1 + e^{\frac{x - x_0}{b}}}$$

where,

$$y_0 = \text{minimum y} - \text{value}$$

$$x_0 = x - \text{value corresponding to peak y}$$

$$a = 50\% \quad \text{point}$$

$$b = \text{slope}$$

A repeated-measures analysis-of-variance (ANOVA) was performed on both within- and between-subjects factors to examine the main effects for each condition and stimulus. Chance performance was considered to be 33% identification for consonants. If an interaction was found between any of the main effects, a simple effects analysis was carried out followed by planned comparisons with a Bonferroni correction between the conditions or stimuli in question. A significant effect in a Bonferroni analysis was calculated by dividing the p-value of 0.05 by n-1 where n was the number of conditions or stimuli in question.

III. RESULTS

A. Categorical perception

1. Normal-hearing listeners—Since no main effect was found for the formant-transition-duration factor, the data for all conditions were averaged across 20- and 40-ms stimuli. The left panels in Figure 1 show perceptual labeling (% identification) by young normal-hearing listeners, who listened to the unprocessed auditory continuum along /ba/-/da/-/ga/. The top panel shows the results for the A condition, while the bottom three panels show results for the auditory continuum simultaneously presented with a visual /ba/ (2nd panel), a visual /da/ (3rd panel), or a visual /ga/ stimulus (bottom panel). Open circles, filled squares, and open triangles represent the subjects' percent identification for /ba/, /da/, and /ga/, respectively. The curves represent the best fit of the 4-parameter sigmoidal function to the data. The left vertical dashed line in each panel represents the estimated phonemic boundary between /ba/ and /da/, while the right dashed line represents the boundary between /da/ and /ga/. The asterisk symbol in the visual /ga/ condition (bottom panel) placed above the /da/ response (filled square) at the token 1 location represents one of the commonly observed McGurk effects (auditory /ba/ + visual /ga/ = perceived /da/).

The A condition (left-top panel) shows the classical pattern of categorical perception, with a significant main effect being observed for the three responses [$F_{(2,12)}$ = 5.2; p<0.05]. Post-hoc analysis with a Bonferroni correction revealed that tokens 1-3 were primarily labeled as /ba/ [$F_{(10,4)}$ = 1805.3; p<0.017], tokens 4-8 labeled as /da/ [$F_{(10,4)}$ =551.7; p<0.017], and tokens 9-11 labeled as /ga/ [$F_{(10,4)}$ = 963.3; p<0.017]. The estimated /ba-da/ boundary was located at about token 4 and the /da-ga/ boundary was at token 8.

The simultaneous presentation of the visual stimulus (bottom 3 panels) generally increased the subjects' overall percent identification towards the visual stimulus while decreasing the response to the incompatible auditory stimuli. For example, the visual /ba/ stimulus (2nd panel) increased the response to /ba/ from 0% to about 40% for tokens between 5 and 11, whereby the peak /da/ response decreased to about 60% and the peak /ga/ response decreased to about 50%. Compared with the effect of the visual /ba/ stimulus, the visual /da/ and /ga/ stimuli produced a similar but slightly smaller effect on the subjects' responses.

The simultaneous presentation of the visual /ba/ stimulus significantly shifted the /ba/-/da/ boundary but not the /da/-/ga/ boundary [$F_{(3,12)} = 11.3$; $p<0.05$]. Compared to the A condition, the visual /ba/ shifted rightward the /ba/-/da/ boundary from the 3.9-token position to the 5-token position, indicating that subjects tended to respond /ba/ to more stimulus tokens when the corresponding visual cue was present. This shift in the /ba/-/da/ boundary was not significant when the A condition was compared to the visual /da/ and /ga/ conditions. Additionally, none of the visual stimuli produced any significant shift for the /da/-/ga/ boundary.

The right panels of Fig. 1 show the results obtained under the same conditions from three elderly listeners. Except for the larger error bars, reflecting the smaller sample size (n=3) than the young normal-hearing listeners (n=14), the elderly listeners produced essentially the same results. For example, the categorical boundary was at about token 4 between /ba/ and /da/, and at about 8 between /da/ and /ga/. Similarly, the visual /ba/ shifted the /ba/-/da/ boundary rightward to about token 5, the visual /da/ and /ga/ slightly shifted the /ba/-/da/ boundary leftward, whereas the visual stimuli had minimal effect on the /da/-/ga/ boundary. This small set of data from elderly listeners suggests that age *per se* plays a negligible role in the present AV tasks (see also Walden *et al*., 1993;Helfer, 1998;Cienkowski and Carney, 2002;Sommers *et al*., 2005).

2. Cochlear-implant listeners—Figure 2 shows perceptual labeling by cochlear-implant listeners (left column). The figure configuration and symbol conventions are the same as Fig. 1. Several differences are apparent between the normal and implant listeners. First, cochlear-implant listeners produced an insignificant main effect for the three responses in the A condition [$F_{(2,6)} = 3.9$; $p>0.05$]. Because of a significant interaction between responses and stimulus tokens [$F_{(20,140)} = 4.2$; $p<0.05$], post-hoc analysis with a Bonferroni correction was conducted to show that only tokens from 1-3 produced significantly higher responses to /ba/ than to /da/ or /ga/ ($p<0.025$). Second, different from the nearly perfect response to the reference tokens by the normal-hearing listeners, the highest /ba/ response obtained by the implant listeners was about 60% to tokens 1-2, followed by 50% /da/ to tokens from 5-8 and 30% /ga/ to tokens 10-11. Although their overall categorical responses were much weaker, the implant listeners produced a proper /ba/-/da/ boundary at token 4 and a slightly rightward shifted /da/-/ga/ boundary at token 9.

Third, a totally different pattern emerged for the visual effect on categorical perception in cochlear-implant listeners than in normal-hearing listeners. Independent of the auditory stimulus, the cochlear-implant listeners showed an almost total bias toward the visual /ba/ cue (2nd panel), and to a lesser extent, toward the visual /da/ (3rd panel) and /ga/ cue (bottom panel). Except for the separate /da/ and /ga/ labeling at high tokens (10 and 11) in the visual /da/ condition, the dominant visual cue wiped out the relatively weak categories that existed in the A condition. On the surface, the McGurk effect (the asterisk symbol in the bottom panel) appeared to be much stronger in the implant listeners (60% labeling to the combined auditory /ba/ and visual /ga/ stimuli) than in the normal listeners (24%). However, note that the overall baseline response to /da/ was also much higher in implant listeners (e.g., 46% to the combined auditory /ga/, i.e., token 11, and visual /ga/ stimuli) than in normal listeners (8%). We shall return to this point in Section C.

3. Cochlear-implant simulations—Figure 2 also shows perceptual labeling by young normal-hearing listeners attending to 4- and 8-channel cochlear-implant simulations (middle and right columns). Like cochlear-implant listeners, normal-hearing listeners presented with cochlear-implant simulations produced relatively weak categorical perception and showed a strong bias toward the visual cue, particularly to the visual /ba/. Moreover, the simulated implant listeners produced a characteristic, albeit relatively weak, categorical boundary for the /ba/-/da/ pair at the token 4 position and /da/-/ga/ boundary close to the token 9 position for the 8-channel simulation condition only. Finally, the simulated implant listeners produced a McGurk effect (~40%, represented by asterisk symbols in the middle and right bottom panels) that was greater than normal-hearing listeners attending to unprocessed stimuli (~20%) but smaller than the actual implant listeners (~60%). Does this mean that actual and simulated implant listeners are more susceptible to the McGurk illusion than normal-hearing listeners? The following sections analyze congruent and incongruent conditions in detail.

B. Congruent AV perception (AV benefit)

Figure 3 shows perceptual labeling results in response to A, V, and congruent AV stimuli in normal-hearing listeners, cochlear-implant listeners, and 4- or 8-channel cochlear-implant simulations. The filled, unfilled, and hatched bars correspond to the percent identification of /b/, /d/, and /g/, respectively. In this case, only the reference tokens were included (i.e., token 1 for /ba/, token 6 for /da/, and token 11 for /ga/).

We also used two measures to define the amount of relative AV benefit that can be derived relative to either the A condition or the V condition. The first relative AV benefit measures the amount of phoneme recognition improvement relative to the A baseline and is defined as (AV-A)/(100-A) with AV and A scores expressed as percent scores (Grant and Seitz, 1998). Similarly, the second AV benefit measures the improvement relative to the V baseline and is defined as (AV-V)/(100-V) with AV and V expressed as percent scores. These relative measures, as opposed to the absolute differences (i.e., AV-A or AV-V), were adopted to avoid potential biases because a wide range of A and V scores occurred in the present diverse group of subjects. For example, a bias occurs because a high A score will certainly produce a low AV benefit score with the absolute measure but not necessarily with the relative measure.

1. Normal-hearing listeners—Normal-hearing listeners demonstrated a significant main effect on modality [$F_{(2,12)} = 9.0$; $p<0.05$], showing nearly perfect performance for the identification of the reference stimuli in the A and AV conditions, but lower and more variable performance in the V condition. On average, the proper labeling with the original unprocessed stimuli was 95% for the A condition, and 94% for the AV condition, but only 80% for the V condition.

Normal-hearing listeners also demonstrated a significant main effect on phoneme identification [$F_{(2,12)} = 23.0$; $p<0.05$], showing an averaged labeling of /ba/ 98% of the time, /da/ 88% of the time, and /ga/ 83% of the time. There was a significant interaction between modality and consonant identification [$F_{(4,10)} = 7.0$; $p<0.05$]. The interaction was due to the fact that the normal-hearing listeners confuse the phonemes /da/ and /ga/ only in the V condition. This result is not too surprising because the voiced bilabial stop consonant, /b/, was more visually salient than either /d/ or /g/.

Averaged across all 3 phonemes, the normal-hearing listeners produced an AV benefit score of -0.14 relative to the A baseline and +0.71 relative to the V baseline. The negative score was due to the fact that the 94% AV percent score was slightly lower the 95% A-alone score. These relative AV benefit scores suggest that the relative signal strength in the auditory domain is strong, producing a ceiling effect.

2. Cochlear-implant listeners—Different from the normal control, the implant listeners produced much lower overall percent correct scores in all conditions. Average performance collapsed across all conditions was 67% for the AV condition, 51% for the A condition, and 52% for the V condition. A significant main effect was found for modality [$F_{(2,6)} = 15.7$; $p<0.05$] in cochlear-implant listeners. A significant main effect was also found for phoneme identification [$F_{(2,6)} = 15.7$; $p<0.05$]. Average performance collapsed across all modalities was significantly higher for /b/ (80%), but no difference was shown between /d/ (44%) and /g/ (46%). A significant interaction was found between modality and phoneme identification [$F_{(4,4)} = 41.5$; $p<0.05$]. This interaction was attributed to the larger difference in performance between /b/ and /d/ in AV (43 percentage points) and V (52 percentage points) conditions, as compared to the auditory only condition (11 percentage points).

The present cochlear-implant listeners were able to integrate the auditory and visual cues to significantly improve the A or the V performance by 15-16 percentage points. The corresponding relative AV benefit score was 0.30 relative to the A baseline and 0.29 relative to the V baseline.

3. Cochlear-implant simulations—First, a significant main effect was found for modality [$F_{(2,5)} = 22.8$; $p<0.05$]. Like actual implant listeners, the simulated implant listeners appeared to benefit from the additional lipreading cue when presented with the degraded auditory stimuli. The performance was improved from 48% with A stimuli to 70% with AV stimuli in the 4-channel condition, and from 63% to 74% in the 8-channel condition. However, unlike actual implant listeners, the V condition produced the highest performance of about 80% in both cochlear-implant simulations, suggesting no integration between the auditory and visual cues.

Second, a significant main effect was found for number of channels [$F_{(1,6)} = 37.5$; $p<0.05$], with the 8-channel condition producing 72% performance and the 4-channel condition producing significantly lower performance at 66%. A significant interaction between channel and modality was also observed [$F_{(2,5)} = 15.6$; $p<0.05$], reflecting lower A performance for the 4-channel condition than the 8-channel condition. This result was expected because the greater spectral resolution associated with the 8-channel would produce better A performance than the 4-channel condition.

Third, a significant main effect was found for phoneme identification [$F_{(2,5)} = 24.3$; $p<0.05$]. In the 4-channel condition, the mean performance was 83%, 64% and 52% for /b/, /d/ and /g/, respectively. In the 8-channel condition, the mean performance was 91%, 62%, and 64% for /b/, /d/, and /g/, respectively. Overall, the simulated implant listeners performed poorer than the normal-hearing listeners when listening to the unprocessed stimuli, but more similarly to the actual cochlear-implant listeners.

Finally, the simulated 4-channel implant listeners produced the following percent correct scores: 50% for A, 81% for V, and 68% for AV; the simulated 8-channel implant listeners produced the following percent correct scores: 63% for A, 81% for V, and 74% for AV. The corresponding relative AV benefit score was 0.36 relative to the A baseline and -0.70 relative to the V baseline in the 4-channel simulation, and was 0.30 and -0.37, respectively, in the 8-channel simulation. Similar to the actual implant listeners, the simulated implant listeners produced similar auditory-only percent scores and AV benefit scores relative to the auditory baseline. Different from the actual implant listeners, the simulated listeners produced higher visual-only scores and negative AV benefit scores relative to the visual baseline.

C. Incongruent AV perception (McGurk effect)

Figure 4 shows perceptual labeling in response to the incongruent condition (the reference auditory /ba/ paired with the visual /ga/) in normal-hearing listeners listening to the unprocessed auditory stimuli, cochlear-implant listeners, 4-channel, and 8-channel cochlear-implant simulations. The filled, unfilled, and hatched bars represent the percentage of the /ba/, /da/, and /ga/ responses, respectively. While a /ba/ response indicates a bias towards the auditory cue and a /ga/ response indicates a bias towards the visual cue, a /da/ response represents integration of the auditory and visual cues, also known as the McGurk effect in this case.

We used two different measures to estimate the size of the McGurk effect. First, hearing-impaired listeners are prone to numerous auditory errors, with some of the errors being consistent with McGurk-like responses, making it difficult to distinguish between these auditory errors and a true McGurk effect (Grant and Seitz, 1998). To overcome this auditory error problem, we followed Grant and Seitz's corrective method (1998) and adjusted the size of the McGurk effect by subtracting the subject's /da/ response to visual /ga/ and auditory /ba/ (i.e., the solid square data point above token 1 in bottom panels of Figs. 1 and 2) from the subject's averaged /da/ response to the auditory /ba/ token (i.e., the solid square data point above token 1 in top panels of Figs. 1 and 2) and the auditory /ga/ token (i.e., the solid square data point above token 11 in top panels of Figs. 1 and 2). We will refer this measure as the error-adjusted McGurk effect.

Second, there may be a general subject response bias. If the bias happens to be a McGurk-type response, it will inflate the size of the McGurk effect. To overcome this response bias problem, we adjusted the size of McGurk effect by subtracting the subject's /da/ response to the incongruent stimulus from the subject's /da/ response to the congruent /ga/ stimulus (i.e., the solid square data point above token 11 in the bottom panels of Figs. 1 and 2). We will refer to this measure as the bias-adjusted McGurk effect.

1. Normal-hearing listeners—A significant effect was found for consonant identification [$F_{(2,12)}$ = 6.5; p<0.05]. The normal-hearing listeners responded 63% of the time to /ba/, 24% to /da/ and 13% to /ga/, implicating a strong auditory bias. The 24% McGurk effect was relatively weak, possibly due to the weak signal strength of the present synthetic stimuli. Interestingly, 6 out of 14 normal-hearing listeners did not experience any McGurk effect, responding to the auditory /ba/ cue 100% of the time. The remaining 8 subjects who experienced the McGurk effect had a mean /da/ response of 42% (sd=23, with a range from 15% to 85%). The unadjusted 24% McGurk effect was significantly higher than the averaged 3% /da/ response to the auditory /ba/ and /ga/ tokens (paired-t test, p<0.05), resulting in an error-adjusted McGurk effect of 21%. On the other hand, the normal subjects produced an 8% /da/ response to the congruent /ga/ token, which was significant lower than the 24% unadjusted McGurk effect (paired-t test, p<0.05). The bias-adjusted McGurk effect size is therefore 16%.

2. Cochlear-implant listeners—A significant effect was found for consonant identification [$F_{(2,6)}$ = 54.1; p<0.05]. The implant listeners responded 4% of the time to /ba/, 60% to /da/ and 36% to /ga/ with the incongruent auditory /ba/ and visual /ga/ stimulus. The implant response pattern was significantly different from the normal control pattern [$F_{(2,19)}$ = 8.6; p<0.05]. For example, the response to the auditory /ba/ cue was greatly reduced from 63% in the normal control to 4% in the implant listeners, suggesting that the auditory signal strength via a cochlear implant was much weaker than the visual signal strength.

On the surface, the McGurk effect appeared to be much stronger in implant listeners, producing a fused /da/ response 60% of the time, compared with the /da/ response 24% of the time in normal control. The 33% error-adjusted McGurk effect was significant (paired t-test, p<0.05) but the 14% bias-adjusted McGurk effect just missed the pre-defined significance criterion (paired t-test, p=0.08). Overall, the 33% error-adjusted and the 14% bias-adjusted McGurk effects in implant subjects were statistically indistinguishable from their 21% and 16% counterparts in the normal control (t-test with two-sample unequal variance; p>0.2). Overall, these adjusted measures suggest that although there is some evidence for integration between incongruent auditory and visual cues, but this integration is weak, if present at all, in post-lingually deafened, adult implant listeners.

3. Cochlear-implant simulations—No significant main effect was found for consonant identification for either the 4-channel [$F_{(2,5)} = 0.9$; p>0.05] or 8-channel [$F_{(2,5)} = 1.0$; p>0.05] condition. In the 4-channel condition, the response was the least to the auditory /ba/ cue (19%), but was increasingly biased toward the fused /da/ cue (36%) and the visual /ga/ cue (45%). In the 8-channel condition, the response was evenly distributed across the auditory /ba/, the fused /da/, and the visual /ga/ cues at near chance performance. Neither adjusted McGurk effect showed any indication of integration between the incongruent auditory and visual cues in these simulated implant listeners.

IV. DISCUSSION

A. Categorical perception

The present results show sharp categories in young and elderly normal-hearing listeners, but greatly weakened categories in actual and simulated cochlear-implant listeners. This finding is consistent with previous studies on hearing-impaired listeners (Walden *et al*., 1990) and pediatric cochlear-implant users (Clark, 2003) who also showed broader than normal boundaries in a similar categorical perception task. Together, these data suggest that categorical perception is affected by hearing loss or reduced spectral resolution, but not by age.

B. Congruent auditory and visual benefit

The present AV benefit in normal-hearing subjects was confounded by the ceiling effect. The actual implant data showed similar 0.30 AV benefit scores relative to either the auditory or the visual baseline. This relative AV benefit score was about half of the 0.67 AV benefit score obtained by hearing-impaired listeners in the Grant and Seitz study (1998). This discrepancy might be due to either the specific and limited set of stimuli, or the perceptual difference between hearing-impaired and cochlear-implant subjects, or both. The simulated implant data suggest that current acoustic simulation models of the cochlear implant (Shannon *et al*., 1995) can simulate auditory perception of degraded acoustic stimuli but cannot simulate the V and AV perception in actual cochlear-implant users.

C. Incongruent auditory and visual integration

When incongruent acoustic and optic cues are present, listeners may be biased toward either auditory or visual cues. Consistent with previous studies (Easton and Basala, 1982; Massaro and Cohen, 1983; Bernstein *et al*., 2000; Cienkowski and Carney, 2002; Clark, 2003; Schorr *et al*., 2005), the present results show that normal-hearing listeners were biased toward the auditory cue (i.e., greater /ba/ response in Fig. 4) while cochlear-implant listeners were biased toward the visual cue (i.e., greater /da/ and /ga/ response in Fig. 4).

This bias appears to depend on the relative signal strength between acoustic and optical stimuli and is likely to influence the degree of integration (e.g., Green *et al*., 1991; Green and Norrix, 1997). In the normal-hearing subjects, the optical signal strength from the animated “Baldi” was relatively weaker than the synthetic acoustic signal strength. In actual and simulated implant listeners, however, the same optical signal became relatively stronger than the improvised acoustic signal. While the unadjusted McGurk effect was much greater in implant subjects than normal subjects, this difference could be totally accounted for by the McGurk-like auditory error responses and response biases (Grant and Seitz, 1998). Overall, the present data showed that the adjusted McGurk effect was weak in normal-hearing and cochlear-implant subjects but totally absent in simulated-implant subjects.

D. Experience and performance

Typically before receiving their devices, post-lingually deafened adult-implant users experience a period of deafness from several months to decades and during which they rely on lipreading. After implantation, users usually need several months to years to achieve asymptotic performance (e.g., Tyler and Summerfield, 1996; Skinner *et al*., 2002). This unique experience may allow them to use the visual cue and integrate the auditory and visual cues to a greater extent than the normal-hearing listeners (Schorr *et al*., 2005; Rouger *et al*., 2007).

Recent brain imaging studies showed strong association between cortical plasticity and cochlear-implant performance: In general, good performance was correlated with the amount of overall cortical activation, not only in the auditory cortex but also in the visual cortex when using the implant only (Giraud *et al*., 2001a; Lee *et al*., 2001; Green *et al*., 2005). Because good performance is correlated with the duration of implant experience, we would expect that both variables are correlated with AV integration. On the other hand, Doucet et al.(2006) found an intramodal reorganization in good implant performers but a profound cross-modality reorganization in poor performers, suggesting that the duration of deafness is correlated with the AV integration.

To address this experience and performance issue, we performed correlational analysis between 2 implant variables (duration of deafness and duration of implant usage) and 7 implant performance measures (Table 3). The implant performance included 3 direct measures in response to A, V, and AV stimuli and 4 derived measures: the AV benefit score relative to A baseline (AV_A), the AV benefit score relative to V baseline (AV_V), the error-adjusted McGurk effect (M_error), and the bias-adjusted McGurk effect (M_bais).

Consistent with previous studies (e.g., Blamey *et al*., 1996; van Dijk *et al*., 1999; Gomaa *et al*., 2003), the duration of deafness is negatively correlated with the duration of the implant usage and the auditory-only performance. However, we found that the duration of deafness is not correlated with any other measures, including the visual-only performance. On the other hand, we found that the duration of implant experience is directly correlated with the A and AV performance, as well as the AV benefit relative to the visual baseline and the McGurk effect. Because visual-only performance is not correlated to any implant performance, the present data suggest, at least in the present post-lingually deafened implant users, that implant experience, rather than auditory deprivation, determines the implant users' ability to integrate both congruent AV cues (i.e., the AV benefit) and incongruent AV cues (i.e., the McGurk effect).

V. CONCLUSIONS

Auditory-only, visual-only, and AV speech perception was conducted in normal-hearing listeners, post-lingually deafened actual implant users, and acoustically-simulated cochlear-implant listeners. Given the limitations of using the synthetic acoustic stimuli, the animated face, and the 3-alternative, forced-choice method, we can reach the following conclusions:

1. Normal-hearing listeners show not only sharp categorical perception of place of articulation but also strong reliance on the auditory cue;
2. Cochlear-implant listeners show weak categorical perception of place of articulation but strong reliance on the visual cue;

3. The implant listeners can derive significant AV benefit from the congruent acoustic and optical cues;
4. Both normal and implant listeners produced a relatively weak McGurk effect in response to the incongruent acoustic and optical cues, possibly due to the weak signal strength in the synthetic stimuli;
5. It is the duration of the implant experience, rather than the duration of deafness, that correlates with the amount of AV integration;
6. The present acoustic simulation model of the cochlear implant can predict the actual implant auditory-only performance but not the AV speech integration.

VI. ACKNOWLEDGEMENTS

We are very grateful for the time and dedication that our cochlear-implant and normal-hearing listeners have offered for this study. We thank Dominic Massaro, Jacques de Villiers, and Paul Hosom at the Center for Spoken Language Understanding, UC Santa Cruz, for their help and for the use of their BALDI software. We also thank Abby Copeland, Jeff Carroll, Hongbin Chen, Sara Ghazi, Jeremy Liu, Sheng Liu, and Kaibao Nie for their assistance in technical support and helpful comments on the manuscript. This work was supported in part by a grant from the National Institutes of Health (RO1 DC002267).

REFERENCES

Barnard, E.; Bertrant, J.; Rundlem, B.; Cole, R.; al., e. 2000. http://cslu.cse.ogi.edu/, Oregon Graduate Institute of Science and Technology (October 16, 2007)

Bergeson TR, Pisoni DB, Davis RA. Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. Ear Hear 2005;26:149–164. [PubMed: 15809542]

Bernstein LE, Auer ET Jr. Moore JK, Ponton CW, Don M, Singh M. Visual speech perception without primary auditory cortex activation. Neuroreport 2002;13:311–315. [PubMed: 11930129]

Bernstein LE, Demorest ME, Tucker PE. Speech perception without hearing. Perception & Psychophysics 2000;62:233–252. [PubMed: 10723205]

Binnie CA, Montgomery AA, Jackson PL. Auditory and visual contributions to the perception of consonants. Journal of Speech and Hearing Research 1974;17:619–630. [PubMed: 4444283]

Blamey P, Arndt P, Bergeron F, Bredberg G, Brimacombe J, Facer G, Larky J, Lindstrom B, Nedzelski J, Peterson A, Shipp D, Staller S, Whitford L. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. Audiol Neurootol 1996;1:293–306. [PubMed: 9390810]

Braida LD. Crossmodal integration in the identification of consonant segments. Q J Exp Psychol A 1991;43:647–677. [PubMed: 1775661]

Bunnell, H. 1996. http://wagstaff.asel.udel.edu/speech/tutorials/synthesis/, Speech Research Laboratory, A.I. duPont Hospital for Children (October 16, 2007)

Burnham D, Dodd B. AV speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. Dev Psychobiol 2004;45:204–220. [PubMed: 15549685]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. Science 1997;276:593–596. [PubMed: 9110978]

Calvert GA, Campbell R. Reading speech from still and moving faces: the neural substrates of visible speech. J Cogn Neurosci 2003;15:57–70. [PubMed: 12590843]

Cienkowski KM, Carney AE. AV speech perception and aging. Ear Hear 2002;23:439–449. [PubMed: 12411777]

Clark G. Cochlear implants in children: safety as well as speech and language. Int J Pediatr Otorhinolaryngol 2003;67(Suppl 1):S7–20. [PubMed: 14662167]

De Gelder B, Bertelson P. Multisensory integration, perception and ecological validity. Trends Cogn Sci 2003;7:460–467. [PubMed: 14550494]

Dodd B. The role of vision in the perception of speech. Perception 1977;6:31–40. [PubMed: 840618]

Doucet ME, Bergeron F, Lassonde M, Ferron P, Lepore F. Cross-modal reorganization and speech perception in cochlear implant users. Brain 2006;129:3376–3383. [PubMed: 17003067]

Easton RD, Basala M. Perceptual dominance during lipreading. Perception & Psychophysics 1982;32:562–570. [PubMed: 7167355]

Erber NP. Auditory, visual, and AV recognition of consonants by children with normal and impaired hearing. Journal of Speech and Hearing Research 1972;15:413–422. [PubMed: 5047880]

Fisher CG. Confusions among visually perceived consonants. Journal of Speech and Hearing Research 1968;11:796–804. [PubMed: 5719234]

Giraud AL, Price CJ, Graham JM, Frackowiak RS. Functional plasticity of language-related brain areas after cochlear implantation. Brain 2001a;124:1307–1316. [PubMed: 11408326]

Giraud AL, Price CJ, Graham JM, Truy E, Frackowiak RS. Cross-modal plasticity underpins language recovery after cochlear implantation. Neuron 2001b;30:657–663. [PubMed: 11430800]

Goh WD, Pisoni DB, Kirk KI, Remez RE. Audio-visual perception of sinewave speech in an adult cochlear implant user: a case study. Ear Hear 2001;22:412–419. [PubMed: 11605948]

Gomaa NA, Rubinstein JT, Lowder MW, Tyler RS, Gantz BJ. Residual speech perception and cochlear implant performance in postlingually deafened adults. Ear Hear 2003;24:539–544. [PubMed: 14663353]

Gordon-Salant S, Fitzgibbons PJ. Selected cognitive factors and speech recognition performance among young and elderly listeners. J Speech Lang Hear Res 1997;40:423–431. [PubMed: 9130210]

Grant KW. Encoding voice pitch for profoundly hearing-impaired listeners. J Acoust Soc Am 1987;82:423–432. [PubMed: 3624647]

Grant KW, Braida LD. Evaluating the articulation index for AV input. J Acoust Soc Am 1991;89:2952–2960. [PubMed: 1918633]

Grant KW, Braida LD, Renn RJ. Single band amplitude envelope cues as an aid to speechreading. Q J Exp Psychol A 1991;43:621–645. [PubMed: 1775660]

Grant KW, Braida LD, Renn RJ. Auditory supplements to speechreading: combining amplitude envelope cues from different spectral regions of speech. J Acoust Soc Am 1994;95:1065–1073. [PubMed: 8132900]

Grant KW, Seitz PF. Measures of AV integration in nonsense syllables and sentences. J Acoust Soc Am 1998;104:2438–2450. [PubMed: 10491705]

Grant KW, Walden BE. Evaluating the articulation index for AV consonant recognition. J Acoust Soc Am 1996;100:2415–2424. [PubMed: 8865647]

Grant KW, Walden BE, Seitz PFS. AV speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and AV integration. The Journal of the Acoustical Society of America 1998;103:2677–2690. [PubMed: 9604361]

Green KM, Julyan PJ, Hastings DL, Ramsden RT. Auditory cortical activation and speech perception in cochlear implant users: effects of implant experience and duration of deafness. Hear Res 2005;205:184–192. [PubMed: 15953527]

Green KP, Kuhl PK, Meltzoff AN, Stevens EB. Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect. Percept Psychophys 1991;50:524–536. [PubMed: 1780200]

Green KP, Norrix LW. Acoustic cues to place of articulation and the McGurk Effect: the role of release bursts, aspiration, and formant transitions. Journal of Speech, Language, and Hearing Research 1997;40:646–665.

Greenwood D. A cochlear frequency-position function for several species - 29 years later. The Journal of the Acoustical Society of America 1990;87:2592–2605. [PubMed: 2373794]

Hay-McCutcheon MJ, Pisoni DB, Kirk KI. Audiovisual speech perception in elderly cochlear implant recipients. Laryngoscope 2005;115:1887–1894. [PubMed: 16222216]

Helfer KS. Auditory and AV recognition of clear and conversational speech by older adults. J Am Acad Audiol 1998;9:234–242. [PubMed: 9644622]

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. J Acoust Soc Am 1995;97:3099–3111. [PubMed: 7759650]

Humes LE. Factors underlying the speech-recognition performance of elderly hearing-aid wearers. J Acoust Soc Am 2002;112:1112–1132. [PubMed: 12243159]

Klatt DH. Software for a cascade/parallel formant synthesizer. The Journal of the Acoustical Society of America 1980;67:971–995.

Kong YY, Stickney GS, Zeng FG. Speech and melody recognition in binaurally combined acoustic and electric hearing. J Acoust Soc Am 2005;117:1351–1361. [PubMed: 15807023]

Lachs L, Pisoni DB, Kirk KI. Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: a first report. Ear Hear 2001;22:236–251. [PubMed: 11409859]

Lee DS, Lee JS, Oh SH, Kim SK, Kim JW, Chung JK, Lee MC, Kim CS. Cross-modal plasticity and cochlear implants. Nature 2001;409:149–150. [PubMed: 11196628]

Massaro, D.; Cohen, M.; Beskow, J.; Cole, R. Developing and evaluating Conversational Agents. In: Chassell, J.; Sullivan, J.; Prevost, S.; Churchill, E., editors. Embodied conversational agents. MIT Press; Cambridge, MA: 2000. p. 287-318.

Massaro, DW. Perceiving Talking Faces: From Speech Perception to a Behavioral Principle. MIT Press; Cambridge, MA: 1998.

Massaro DW, Cohen MM. Evaluation and Integration of Visual and Auditory Information in Speech Perception. Journal of Experimental Psychology: Human Perception and Performance 1983;9:753–771. [PubMed: 6227688]

Massaro DW, Cohen MM. Tests of AV integration efficiency within the framework of the fuzzy logical model of perception. J Acoust Soc Am 2000;108:784–789. [PubMed: 10955645]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976;264:746–748. [PubMed: 1012311]

Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. The Journal of the Acoustical Society of America 1955;27:338–352.

Montgomery AA, Walden BE, Schwartz DM, Prosek RA. Training AV speech reception in adults with moderate sensorineural hearing loss. Ear Hear 1984;5:30–36. [PubMed: 6706024]

Moody-Antonio S, Takayanagi S, Masuda A, Auer ET Jr. Fisher L, Bernstein LE. Improved speech perception in adult congenitally deafened cochlear implant recipients. Otol Neurotol 2005;26:649–654. [PubMed: 16015162]

Munson B, Donaldson GS, Allen SL, Collison EA, Nelson DA. Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability. The Journal of the Acoustical Society of America 2003;113:925–935. [PubMed: 12597186]

Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, Howard MA, Brugge JF. AV processing represented in the human superior temporal gyrus. Neuroscience 2007;145:162–184. [PubMed: 17241747]

Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 1992;336:367–373. [PubMed: 1354376]

Rosen SM, Fourcin AJ, Moore BC. Voice pitch as an aid to lipreading. Nature 1981;291:150–152. [PubMed: 7231534]

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb Cortex 2007;17:1147–1153. [PubMed: 16785256]

Rouger J, Lagleyre S, Fraysse B, Deneve S, Deguine O, Barone P. Evidence that cochlear-implanted deaf patients are better multisensory integrators. Proc Natl Acad Sci U S A 2007;104:7295–7300. [PubMed: 17404220]

Sams M, Manninen P, Surakka P, Helin P, Katto R. McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. Speech Communication 1998;26:75–87.

Schorr EA, Fox NA, van Wassenhove V, Knudsen EI. AV fusion in speech perception in children with cochlear implants. Proc Natl Acad Sci U S A 2005;102:18748–18750. [PubMed: 16339316]

Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. J Acoust Soc Am 1999;106:L71–74. [PubMed: 10615713]

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–304. [PubMed: 7569981]

Skinner MW, Holden LK, Whitford LA, Plant KL, Psarros C, Holden TA. Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. Ear Hear 2002;23:207–223. [PubMed: 12072613]

Sommers MS, Tye-Murray N, Spehar B. AV speech perception and AV enhancement in normal-hearing younger and older adults. Ear Hear 2005;26:263–275. [PubMed: 15937408]

Stickney GS, Zeng FG, Litovsky RY, Assmann PF. Cochlear implant speech recognition with speech masker. J. Acoust. Soc. Am 2004;116:1081–1091. [PubMed: 15376674]

Sumby W, Pollack I. Visual Contribution to Speech Intelligibility in Noise. J Acoust Soc Am 1954;26:212–215.

Summerfield Q. Use of Visual Information for Phonetic Perception. Phonetica 1979;36:314–331. [PubMed: 523520]

Turner CW, Robb MP. Audibility and recognition of stop consonants in normal and hearing-impaired subjects. The Journal of the Acoustical Society of America 1987;81:1566–1573. [PubMed: 3584694]

Turner CW, Smith SJ, Aldridge PL, Stewart SL. Formant transition duration and speech recognition in normal and hearing-impaired listeners. The Journal of the Acoustical Society of America 1997;101:2822–2825. [PubMed: 9165736]

Tyler RS, Summerfield AQ. Cochlear implantation: relationships with research on auditory deprivation and acclimatization. Ear Hear 1996;17:38S–50S. [PubMed: 8807275]

van Dijk JE, van Olphen AF, Langereis MC, Mens LH, Brokx JP, Smoorenburg GF. Predictors of cochlear implant performance. Audiology 1999;38:109–116. [PubMed: 10206520]

Van Tasell DJ, Soli SD, Kirby VM, Widin GP. Speech waveform envelope cues for consonant recognition. J Acoust Soc Am 1987;82:1152–1161. [PubMed: 3680774]

van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in AV speech perception. Neuropsychologia 2007;45:598–607. [PubMed: 16530232]

Walden BE, Busacco DA, Montgomery AA. Benefit from visual cues in AV speech recognition by middle-aged and elderly persons. J Speech Hear Res 1993;36:431–436. [PubMed: 8487533]

Walden BE, Erdman SA, Montgomery AA, Schwartz DM, Prosek RA. Some effects of training on speech recognition by hearing-impaired adults. J Speech Hear Res 1981;24:207–216. [PubMed: 7265936]

Walden BE, Montgomery AA, Prosek RA, Hawkins DB. Visual biasing of normal and impaired auditory speech perception. J Speech Hear Res 1990;33:163–173. [PubMed: 2314076]

Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, Cao K. Speech recognition with amplitude and frequency modulations. Proc Natl Acad Sci U S A 2005;102:2293–2298. [PubMed: 15677723]

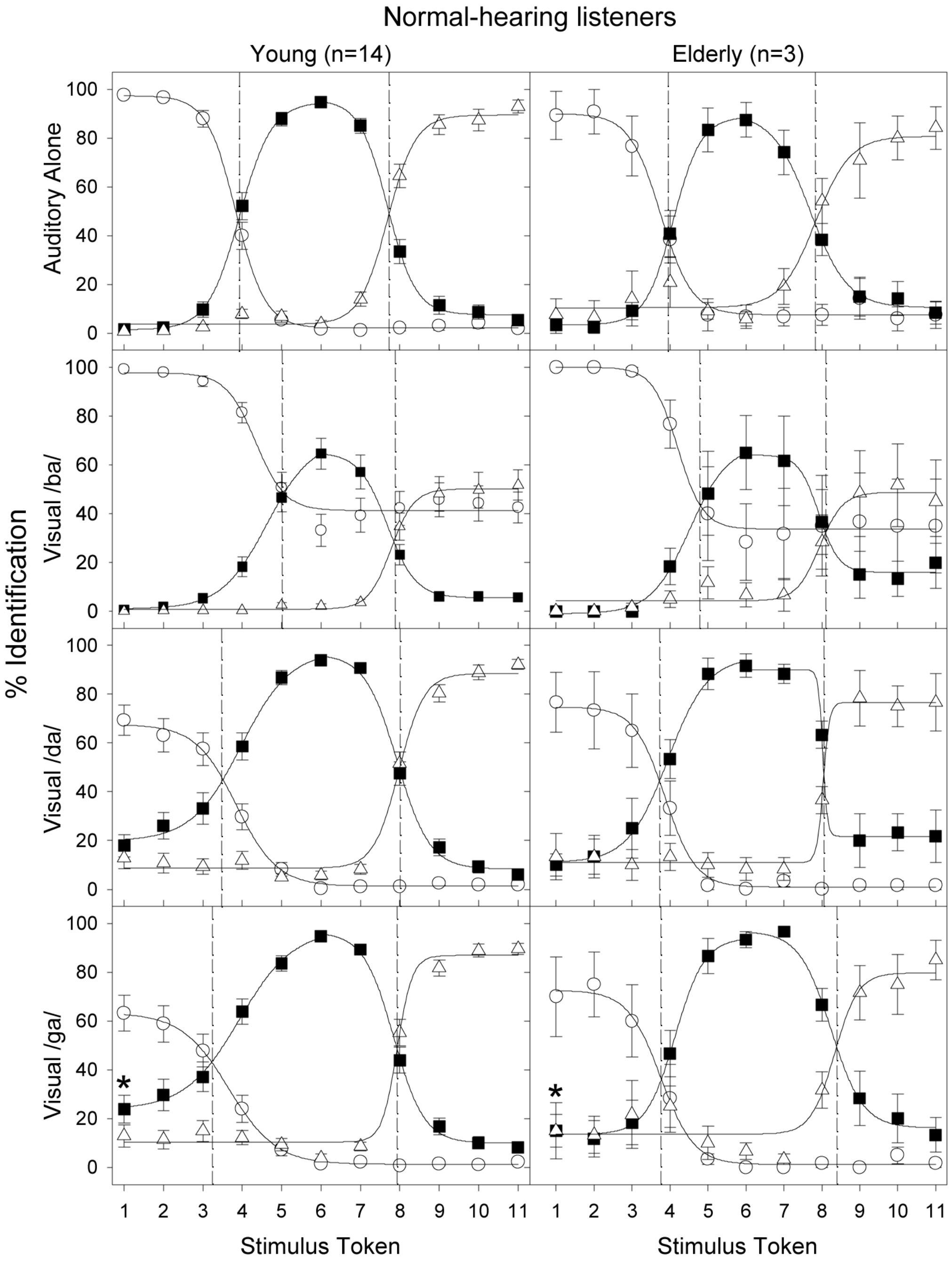

Figure 1.

Percent identification as a function of consonant continua in young (left column) and elderly (right column) normal-hearing listeners. The top panel shows the results for the auditory-alone continuum. The data for the three AV conditions are shown in separate panels (visual /b/: second row; visual /d/: third row; and visual /g/: bottom row). Open circles (○), filled squares (▪), and open triangles (δ) represent the percentage response to /b/, /d/, and /g/, respectively. Error bars represent the standard error of the mean. Sigmoidal 4-parameter functions were fitted to the data to reveal /b/-/d/ and /d/-/g/ boundaries. Vertical dashed lines show where these boundaries occur along the continuum. An asterisk (*) denotes one of the commonly observed

McGurk effects, i.e., when subjects responded /da/ when a visual /ga/ face was paired with the reference auditory /ba/ sound (token 1).

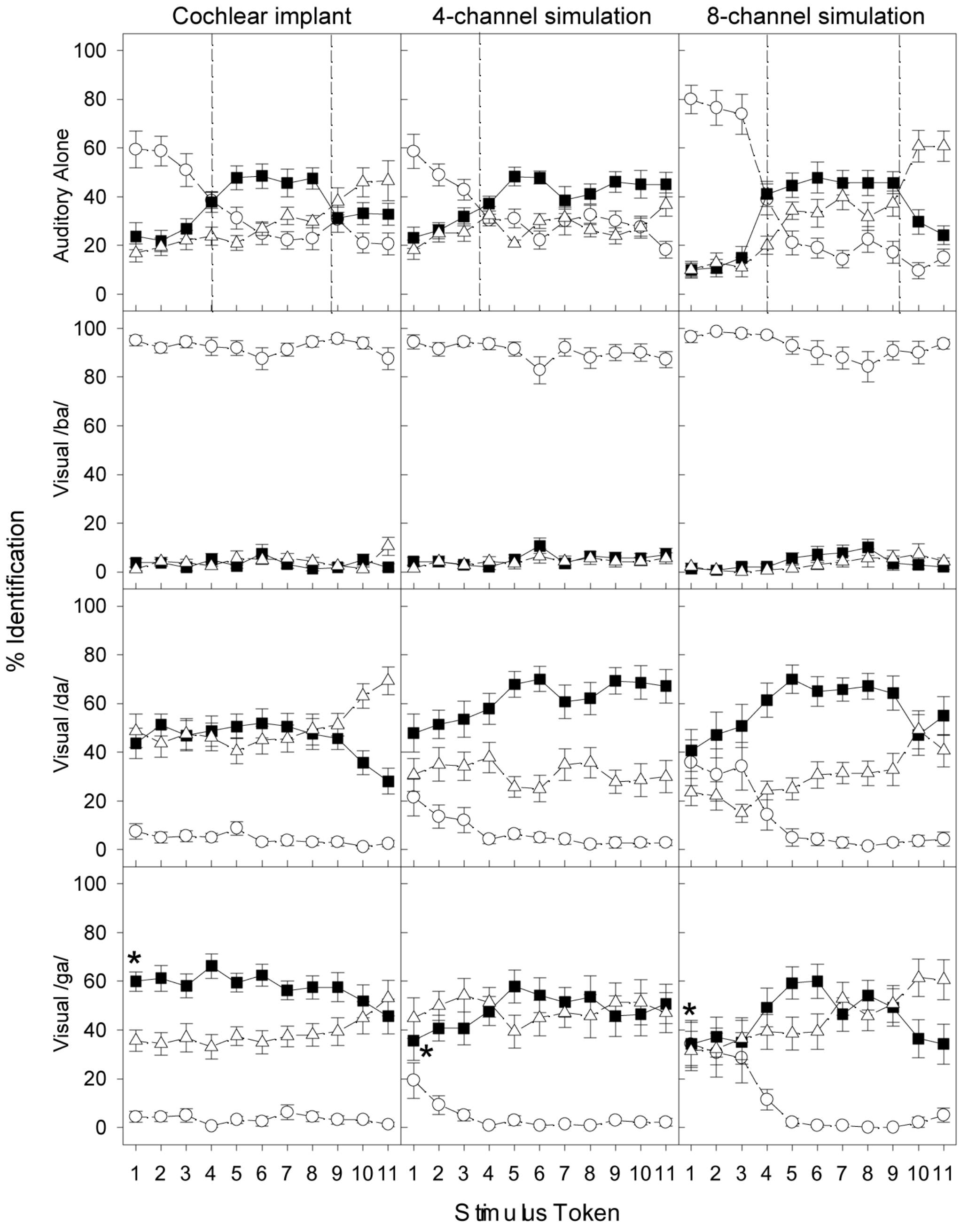

Figure 2.

Percent identification as a function of consonant continuum in the cochlear-implant listeners (first column), and the 4-channel (middle column) and 8-channel (right column) simulated implant listeners. The top panels show the results for the auditory-alone continuum. The data for the three AV conditions are shown in separate panels (visual /b/: second row; visual /d/: third row; and visual /g/: bottom row). Open circles, filled squares, and open triangles represent the percentage response to /b/, /d/, and /g/, respectively. Error bars represent the standard error of the mean. An asterisk (*) denotes the McGurk effect, i.e., when subjects responded /da/ when a visual /ga/ face was paired with the reference auditory /ba/ sound (token 1).

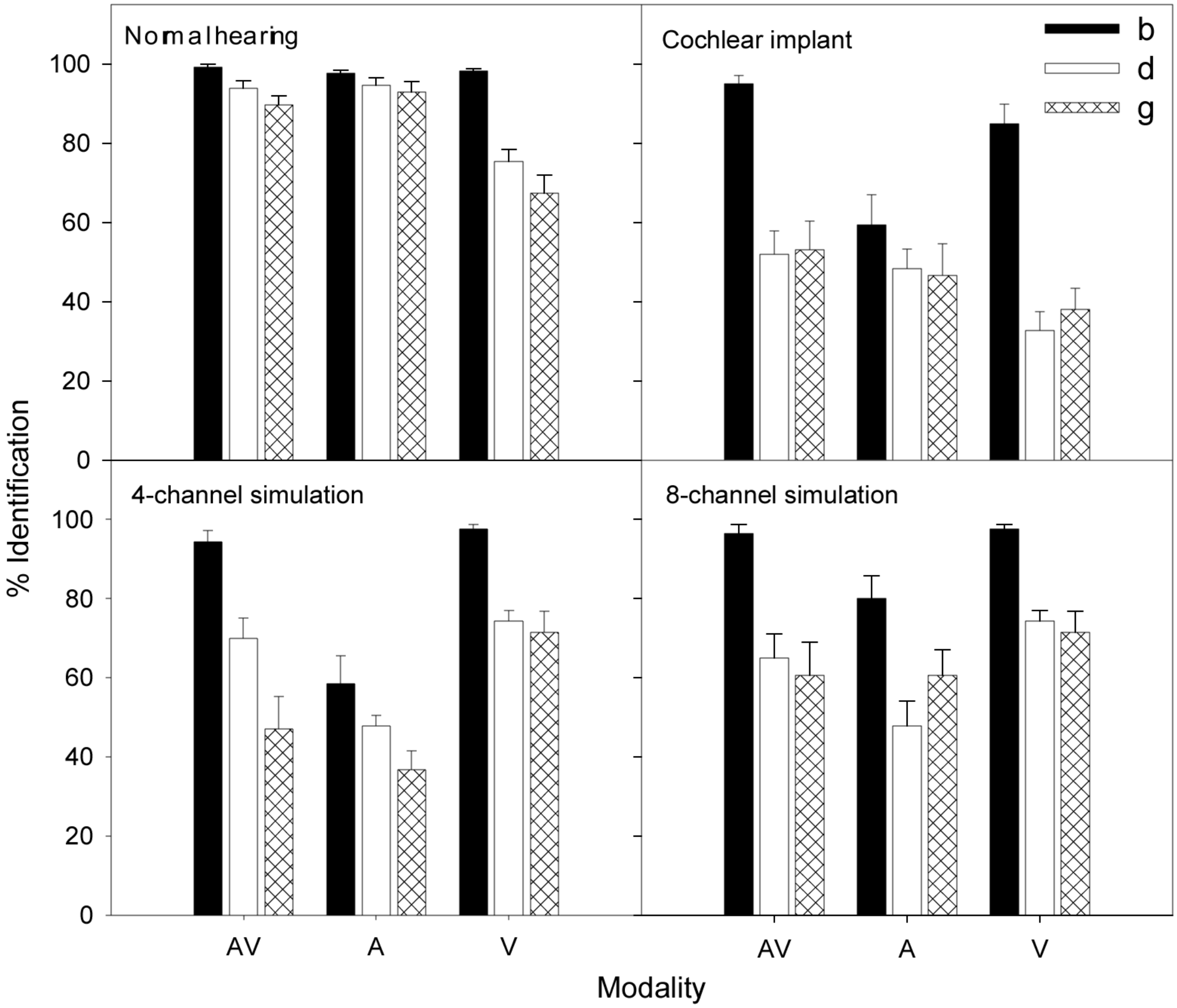

Figure 3.
Performance of all 4 groups of subjects in the congruent AV, A, and V conditions. Error bars represent the standard error of the mean.

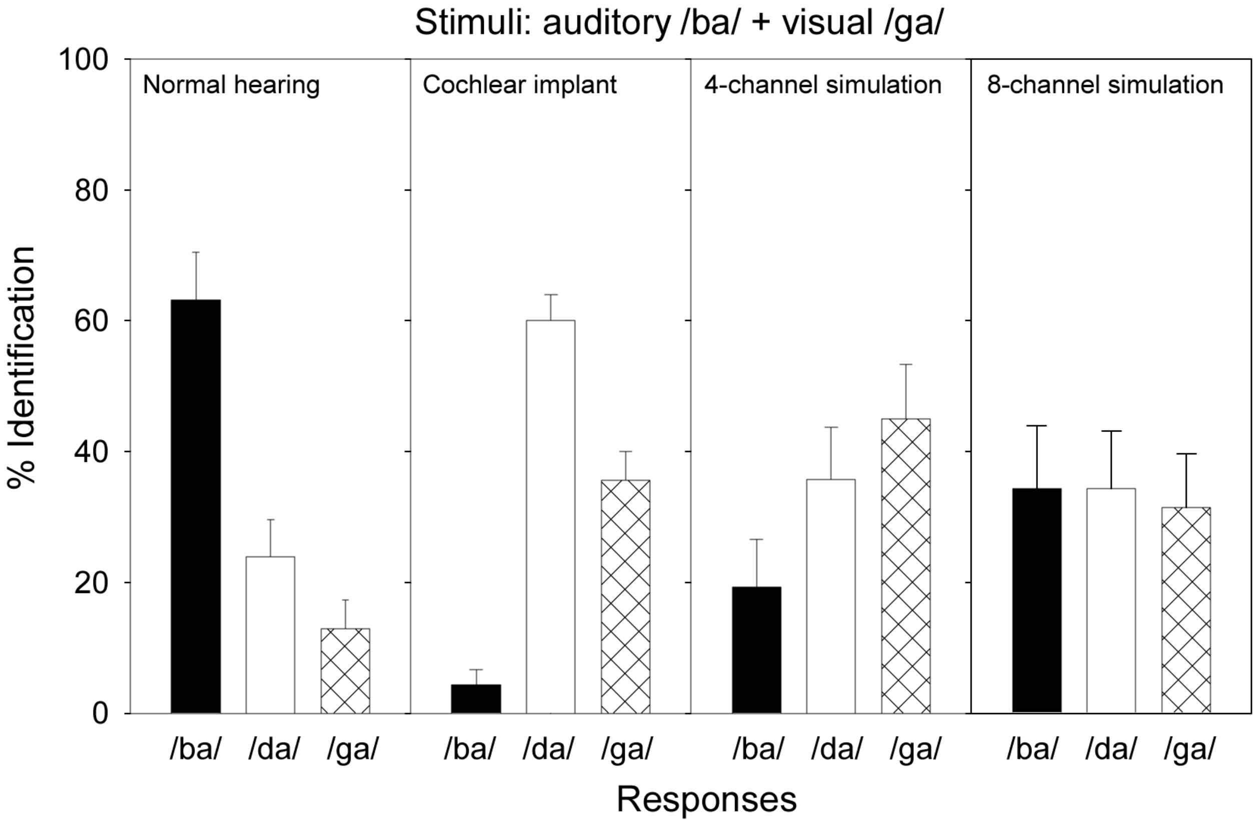

Figure 4.

Percent identification of /ba/, /da/, or /ga/ in all 4 groups of subject for the incongruent AV condition, in which an auditory /ba/ cue was paired with a visual /ga/ cue. Error bars represent the standard error of the mean.

TABLE 1

Demographics of the cochlear-implant listeners, including age, implant type, etiology, duration of deafness, years of experience with the cochlear implant, and percent correct scores in consonant and vowel recognition. Normal-hearing listeners typically score >90% on these consonant and vowel tests (Hillenbrand *et al*., 1995; Shannon *et al*., 1999).

Subject	Age (years)	Implant	Etiology	Duration of Deafness (years)	Implant Experience (years)	% Cons	% Vowels
1	61	Nucleus 22	Genetic	9	5	72	59
2	78	Nucleus 24	Unknown	12	1	44	31
3	70	Nucleus 24	Unknown	11	3	54	51
4	80	Med-El	Genetic	49	1	8	25
5	58	Nucleus 24	Genetic	44	1	12	22
6	46	Nucleus 22	Trauma	1	11	71	79
7	68	Clarion II	Genetic	16	2	59	51
8	69	Nucleus 24	Virus	1	6	77	63

TABLE 2

Starting formant frequencies for each consonant token as well as steady-stat formant frequencies for the vowel (/a/).

Stimulus token	F1 (Hz)	F2 (Hz)	F3 (Hz)
/b/ 1	300	700	2800
2	300	900	2800
3	300	1100	2800
4	300	1300	2800
5	300	1500	2800
/d/ 6	300	1700	2800
7	300	1700	2600
8	300	1700	2400
9	300	1700	2200
10	300	1700	2000
/g/ 11	300	1700	1800
/a/	720	1250	2500

TABLE 3

Correlational analysis between implant variables and implant performance in 8 post-lingually deafened cochlear-implant listeners. The implant variables included the duration of deafness (Deaf) and duration of implant usage (CI). The implant performance included 3 direct measures in response to congruent auditory and visual cues: the auditory-only score (A), the visual-only score (V) the AV score (AV), as well as 4 derived measures (see text for details): the AV benefit score relative to A baseline (AV_A), the AV benefit score relative to V baseline (AV_V), the error-adjusted McGurk effect (M_error), and the bias-adjusted McGurk effect (M_bias). Pearson correlation coefficient was used and significant correlation at the 0.05 level was labeled by the asterisk symbol.

	Deaf	**CI**	**A**	**V**	**AV**	**AV_A**	**AV_V**	**M_error**	**M_bias**
Deaf	1	-0.67*	-0.65*	0.06	-0.38	0.20	-0.41	-0.44	-0.23
CI		1	0.87*	-0.09	0.82*	0.13	0.87*	0.77*	0.63*
A			1	0.02	0.84*	0.14	0.80*	0.90*	0.83*
V				1	0.24	0.48	-0.27	0.41	0.23
AV					1	0.37	0.87*	0.77*	0.89*
AV_A						1	0.17	0.14	-0.11
AV_V							1	0.63*	0.65*
M_error								1	0.87*
M_bias									1