The Benefits of Combining Acoustic and Electric Stimulation for the Recognition of Speech, Voice and Melodies

Michael F. Dorman^a Rene H. Gifford^b Anthony J. Spahr^a Sharon A. McKarns^a

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

^bDepartment of Otorhinolaryngology, Mayo Clinic, Rochester, Minn., USA

Key Words

Cochlear implant • Low-frequency hearing • Electric and acoustic stimulation

Abstract

Fifteen patients fit with a cochlear implant in one ear and a hearing aid in the other ear were presented with tests of speech and melody recognition and voice discrimination under conditions of electric (E) stimulation, acoustic (A) stimulation and combined electric and acoustic stimulation (EAS). When acoustic information was added to electrically stimulated information performance increased by 17–23 percentage points on tests of word and sentence recognition in quiet and sentence recognition in noise. On average, the EAS patients achieved higher scores on CNC words than patients fit with a unilateral cochlear implant. While the best EAS patients did not outperform the best patients fit with a unilateral cochlear implant, proportionally more EAS patients achieved very high scores on tests of speech recognition than unilateral cochlear implant patients.

Copyright © 2007 S. Karger AG, Basel

Introduction

Patients who have been fit with a cochlear implant and who have residual low-frequency hearing in one or both ears can integrate electrically elicited percepts (E) from the implant and acoustically elicited percepts (A) from a hearing aid in the service of speech understanding [e.g., von Ilberg et al., 1999; Ching et al., 2004; Gantz et al., 2005; Kiefer et al., 2005; Kong et al., 2005]. That is to say, speech understanding is better with combined electric and acoustic stimulation (EAS) than with either E-alone or A-alone stimulation. The improvement in speech understanding when both electrically and acoustically elicited information is available is especially noticeable when speech signals are presented in noise. It has been proposed that the improvement in noise is due to the addition of fine-grained information about voice pitch – information not available from a cochlear implant – which enables a listener to take advantage of pitch differences between speakers and noise and to segregate speech targets from noise [e.g., Turner et al., 2004; Chang et al., 2006; Qin and Oxenham, 2006].

We have studied a group of patients who have a fully inserted cochlear implant in one ear and who have low-frequency residual hearing in the other ear. In the non-implanted ear the mean thresholds at 500 Hz and lower

Fig. 1. Mean audiogram for 15 patients with low-frequency residual hearing.

were 53 dB HL and better. Thresholds at 1 kHz and above were 81 dB HL and poorer. All patients used amplification for the ear with residual hearing. In this report we describe the gain in speech recognition that was obtained when patients were allowed to use their residual hearing in addition to their cochlear implant. Information of this type is not novel and our data only add to the growing data set that documents the usefulness of low-frequency auditory information when added to the information delivered by a cochlear implant. What is novel about our study is that we have compared the performance of our patients in EAS test conditions to the performance of 2 groups of patients fit with a conventional cochlear implant. One group was a random sample of conventional implant patients [Helms et al., 1997]. At issue was whether the mean performance of our EAS patients was higher than the mean performance of a random sample of conventional patients. A second control group was composed of above-average patients (50% correct or better on CNC words). At issue was whether the addition of acoustically elicited information elevated performance to a level that is not reached by the highest performing patients who receive only electrically elicited percepts. If this is the case, then it is reasonable to suppose that low-frequency acoustic hearing provides information not available from a conventional cochlear implant.

Methods

Subjects

EAS Patients. The research group consisted of 15 subjects with a fully inserted cochlear implant in one ear and low-frequency hearing in the other ear. Figure 1 displays the mean audiometric thresholds in the nonimplanted ear. The mean thresholds at 0.25, 0.5, 0.75, 1.0, 2.0 and 4.0 kHz were 38, 53, 69, 81, 99 and 104 dB HL, respectively. At the time of testing, all subjects had at least 5 months experience with E stimulation (range: 5 months to 7 years) and at least 5 years of experience with amplification prior to implantation. In accordance with IRB guidelines at Arizona State University, written informed consent was obtained from all subjects. The subjects were paid an hourly wage for their participation.

Conventional Cochlear Implant Patients. The scores for EAS patients were compared to scores from 2 groups of patients who used E stimulation only. One group was composed of 54 patients tested by Helms et al. [1997]. This sample was selected because the CNC scores are representative of a random sample of implant patients. The second group consisted of 65 subjects fit with a unilateral cochlear implant who were tested in our laboratory and who were chosen because they had average (50% correct) or better CNC scores. Twenty-seven patients used the Nucleus Corporation 3G device, 20 used the Advanced Bionics Corporation High Resolution device, 3 used the Med El CIS adaptation for the Ineraid electrode array, and 15 used the Med El CIS Pro Plus device. The patients were originally recruited for a comparative study of device performance [Spahr and Dorman, 2004; Spahr et al., 2007] and included the highest performing patients in the USA. Therefore, this sample is not representative of the complete range of implant performance. Rather, the sample represents above average performance – indeed, the best performance allowed by the current generation of cochlear implants.

Test Materials and Design

Performance was measured in A, E and EAS conditions. Test order was randomized for each subject.

To minimize the contribution of A stimulation to performance in E condition, each subject's hearing aid was removed and the nonimplant ear was occluded with an EAR foam plug and a circumaural headphone. Probe measurements indicated attenuation of 20 dB at 250 Hz and 27.5 dB at 500 Hz with the plug and headphone combination. Total attenuation (foam plug and headphone plus hearing loss) at 250 Hz was 73.5 dB SPL; at 500 Hz attenuation was 89 dB SPL.

The test battery included measures of vowel, consonant, word, sentence and melody recognition in quiet, voice discrimination in quiet and sentence recognition in noise (+10 dB and +5 dB SNR). Signals were presented at 70 dB SPL via a loudspeaker placed at a distance of 1 m in front of the listener. Prior to testing, each subject's hearing aid settings were electroacoustically verified with a simulated speech map fitting system using the Verifit simulated real-ear mode. That is, all measurements were made in the test chamber (with the 2-cc coupler) and were converted to an estimated real ear SPL. Given that all speech stimuli were presented at an overall level of 70 dB SPL, the subjects' hearing aids were only required to meet prescribed targets for the average-level speech signal, which is also presented at 70 dB SPL.

CNC Words. A 50-item CNC word list [Peterson and Lehiste, 1962] was used as test material.

Sentence Material. Lists of 20 sentences each from the multiple-talker, AzBio sentence corpus [Spahr and Dorman, 2004] were used as test material. The sentences were spoken by male and female speakers in a conversational speaking style. The sentences ranged in length from 4 to 7 words. Sentences were scored as words correct. The sentences were presented in quiet, at ± 10 dB SNR and at ± 5 dB SNR. The noise was 4-talker babble from an Auditec CD. The noise started 100 ms before the onset of the signal and ended 100 ms after the end of the signal.

Voice Discrimination. A total of 108 words produced by 5 males and 5 females were drawn from a digital database developed at the Speech Research Laboratory at Indiana University, Bloomington [Clopper et al., 2002]. Patients were presented with pairs of words. Within each condition, half of the pairings were produced by the same talker and half were produced by different talkers. The words in the pairings always differed, e.g., 1 male talker might say 'ball' and the other male talker might say 'brush'. Across the different talker pairs, each talker was paired with every other talker an equal number of times. Participants responded 'same' or 'different' by pressing 1 of 2 buttons. Responses were scored for between-gender (male vs. female) contrasts and for within-gender (male vs. male and female vs. female) contrasts [Kirk et al., 2002].

Melody Recognition. A total of 33 common melodies (e.g., Yankee Doodle, London Bridge) were created for this test. Each melody consisted of 16 equal-duration notes, synthesized with MIDI software that used samples of a grand piano [Hartmann and Johnson, 1991]. The frequencies of the notes ranged from 277 to 622 Hz. The average note was concert A (440 Hz) ± 1 semitone. The melodies were created without distinctive rhythmic information. Prior to testing, patients were asked to select 5 familiar melodies from the list of 33 melodies. After a brief practice period, patients were presented with a melody and asked to identify it by pressing a button from a list containing the 5 preselected melodies. The order of the items was randomized in the test list.

Vowel Recognition without Duration Cues. Using the KLATT software, 13 vowels were created in /bVt/ format ('bait, Bart, bat, beet, Bert, bet, bit, bite, boat, boot, bought, bout, but'). The vowel formats were brief (90 ms) and of equal duration so that vowel length would not be a cue to identity [Dorman et al., 1989]. During a practice session patients heard each vowel presented twice while the word was visually displayed on the computer screen. Patients then completed 2 repetitions of the test procedure, with feedback, as a final practice condition. In the test condition there were 5 repetitions of each stimulus. The order of the items was randomized in the test list.

Consonants in /e/ Environment. Twenty consonants were recorded in 'eCe' format, e.g., 'a bay, a day, a gay', etc. A single male talker made 5 productions of each token. The pitch and vocalic portion of each token was intentionally varied. During a practice session patients heard each signal twice while the word was visually displayed on the computer screen. Patients then completed 2 repetitions of the test procedure, with feedback, as a final practice condition. In the test condition patients heard all 100 tokens (5 productions of each consonant). The order of items was randomized in the test list.

Results

Performance in A, E and EAS Conditions

Repeated-measure ANOVAs were performed to assess the effect of test conditions for the 9 types of stimulus material. Due to the number of tests, α was set to $0.05/n$, where $n = 9$ (i.e., 0.005). Differences among the test conditions (E, A and EAS) were evaluated by the Bonferroni all-pairs test. Percent correct scores as a function of test condition and test material are presented in figures 2 and 3. As shown in figure 2, for CNC words there was a significant effect of test condition: A = 26.5% correct, E = 53.6% correct, EAS = 72.6% correct ($F_{2, 28} = 41.86$, $p < 0.000000$). All scores were significantly different from one another. For consonants, there was a significant effect of test condition: A = 45.6% correct, E = 62.7% correct, EAS = 71.5% correct ($F_{2, 28} = 21.61$, $p < 0.000000$). Posttests indicated that E and EAS scores were higher than A. However, EAS was not higher than E. For vowels there was not a significant main effect of test condition when using α at 0.005: A = 57.6% correct, E = 49.4% correct, EAS = 68.3% correct ($F_{2, 38} = 4.54$, $p = 0.019$).

For sentences in quiet there was a significant effect of test condition: A = 38.7% correct, E = 67.3% correct, EAS = 84.0% correct ($F_{2, 28} = 30.85$, $p < 0.000000$). All scores were significantly different from one another. For sentences at +10 dB SNR there was a significant effect of test condition: A = 24.1% correct, E = 42.6% correct, EAS = 65.2% correct ($F_{2, 28} = 32.05$, $p < 0.000000$). All scores were significantly different from one another. For sentences at +5 dB SNR, there was a significant effect of test condition: A = 9.9% correct, E = 21.8% correct, EAS = 43.8% correct ($F_{2, 28} = 30.85$, $p < 0.000000$). All scores were significantly different from one another.

Scores for melody recognition and within- and between-gender voice discrimination are shown in figure 3. For melody recognition there was a significant effect of condition: A = 70.6% correct, E = 52% correct, EAS = 71.2% correct ($F_{2, 28} = 6.90$, $p = 0.004$). Posttests indicated that E was poorer than A and EAS and that EAS was not different from A. For within-gender voice discrimination there was not a significant difference in performance as a function of test condition: A = 72.4, E = 69.9, EAS = 71.9 ($F_{2, 28} = 1.60$, $p = 0.220$). For between-gender voice discrimination there was not a significant difference in performance as a function of test condition: A = 92.6, E = 87.7, EAS = 94.1, ($F_{2, 28} = 3.28$, $p = 0.052$). For these conditions scores were near the ceiling in performance and this may have obscured differences among conditions.

Fig. 2. Mean scores and standard deviations for E, A and EAS conditions. Significant differences among conditions are indicated, e.g., $E > A$, for each type of test material. **a** CNC words. **b** Consonants. **c** Vowels. **d** Sentences – quiet. **e** Sentences +10 dB SNR. **f** Sentences +5 dB SNR.

Fig. 3. Mean scores and standard deviations for E, A and EAS conditions. Significant differences among conditions are indicated, e.g., $E > A$, for each type of test material. **a** Melodies. **b** Voice – within. **c** Voice – between.

Discussion

As others have reported previously, we find that listeners are able to integrate acoustically and electrically elicited percepts in the service of speech perception, i.e., scores on tests of word and sentence understanding in quiet and sentence understanding in noise were better in EAS conditions than in E conditions. For word under-

standing in quiet, the improvement was 19 percentage points. For sentences in quiet, the improvement was 17 percentage points. For sentences at +10 and +5 dB SNR the improvements were 23 and 22 percentage points, respectively. Thus, we find similar magnitude gains in performance for single words, for sentences in quiet and for sentences in noise.

Other researchers have suggested that the benefits of adding low-frequency acoustic information to electric information may be derived from an improved representation of voice pitch [e.g., Turner et al., 2004; Chang et al., 2006]. A better representation could enable listeners to use voice pitch to separate a target voice from a background of other voices. This is an appealing hypothesis for the often reported improvement for speech understanding in noise in EAS conditions. The hypothesis is supported by several lines of evidence including those from simulations of EAS which show that (i) low-passed speech that contains only the fundamental frequency, and possibly a harmonic, of the voice aids speech understanding in noise [Chang et al., 2006] and (ii) a sine wave modulated at the frequency of the voice aids speech understanding in noise [Brown and Bacon, 2007].

In our study the recognition of words and sentences presented in quiet benefited significantly from adding low-frequency acoustic information to E stimulation. It is not easy to craft an account of how an improved representation of pitch could account for these results. However, a theory of vowel recognition by Miller [1989] offers a possibility. In Miller's theory pitch plays a significant role in vowel recognition by interacting with the frequency of the first formant to form 1 dimension of a 3-dimensional perceptual space. On this view, we might expect that vowel recognition would be significantly improved when A is added to E. However, this was not the case, at least when using our very stringent criterion for significance ($\alpha = 0.005$). On the other hand, performance in the EAS condition was 19 percentage points better than performance in the E condition – a gain in performance equal to that for CNC words.

Pitch may not be the only attribute of low-frequency acoustic information that benefits speech understanding when A is added to E. For example, in both quiet and noise, acoustic information about the frequency of the low, first formant could provide a frequency-appropriate reference against which higher frequency information provided by the cochlear implant could be integrated.

It was not surprising to find that recognition of melodies was better when listeners used their residual low-frequency hearing than when they used their cochlear implant.

It was surprising to find that within-gender voice discrimination, i.e., the ability to discriminate between 2 male or 2 female voices, was as poor with low-frequency acoustic hearing as with a cochlear implant. Speaker identity is specified by both voice pitch and formant frequencies [e.g., Childers and Wu, 1991; Smith et al., 2007].

Fig. 4. Percent correct CNC words for conventional implant patients in Helms et al. [1997] and for the patients in the current study using E and EAS stimulation. Each dot indicates the performance of a single patient. Group mean scores are indicated by a horizontal line.

It is reasonable to suppose that performance was poor when listening via a cochlear implant because neither voice pitch nor formant frequencies were accurately represented. Poor performance when listening via low-frequency acoustic hearing may have been the result of impaired frequency difference limens in the region of elevated thresholds [e.g., Tyler et al., 1983; Peters and Moore, 1992] and inaudible second and higher formant frequencies.

Comparing E and EAS Performance to the Performance of a Random Sample of Conventional Cochlear Implant Patients

In figure 4 we show data for CNC recognition (i) from 54 patients fit with a unilateral cochlear implant [data from Helms et al., 1997], (ii) from our patients in the E condition and (iii) from our patients in the EAS condition. The mean score from the Helms sample was 55% correct and the mean score for our E stimulation condition was 54% correct. The distribution of scores was similar in both data sets. Thus, our E scores are representative of scores from a random sample of conventional cochlear implant patients. Having established that equivalence, we can ask how adding low-frequency acoustic information affects the distribution of scores. As shown in figure 4, the effect is to bring up the bottom of the distribution so that the lowest scores in the EAS condition are at or near the mean of the scores in the E condition.

Fig. 5. Percent correct scores for patients, fit with conventional cochlear implants, who scored 50% correct or better on CNC words (average and above) and for patients in the current study using EAS. Each dot indicates the performance of a single patient. **a** CNC words. **b** Consonants. **c** Vowels. **d** Sentences – quiet. **e** Sentences +10 dB SNR. **f** Sentences +5 dB SNR.

Due to a possible ceiling effect it is not clear whether adding A to E improves the scores at the high end of the distribution. The mean scores improve from 54 to 73% correct. Thus, EAS patients with low-frequency thresholds of 53 dB and better at 500 Hz and 81 dB and poorer at 1–8 kHz achieve, on average, higher scores than patients fit with a conventional cochlear implant.

The mean score (73% correct) in our EAS condition for patients with a full insertion of an electrode in one ear

and low-frequency hearing in the other ear is similar to the mean score of 79% correct reported by Gantz et al. [2005] for patients (i) who receive a relatively short insertion (10 mm) of an electrode array in one ear, (ii) who have preserved low-frequency hearing in that ear and (iii) who have low-frequency hearing in the contralateral ear. Thus, there are at least 2 surgical paths to better speech understanding for patients with bilaterally symmetrical, low-frequency hearing.

Fig. 6. Percent correct scores for patients, fit with conventional cochlear implants, who scored 50% correct or better on CNC words (average and above) and for patients in the current study using EAS. Each dot indicates the performance of a single patient. **a** Melody recognition. **b** Voice – within. **c** Voice – between.

Comparing EAS Performance to the Performance of the Best Conventional Cochlear Implant Patients

At issue in this section is whether EAS performance is better than the very best performance from conventional cochlear implant patients. If so, then low-frequency acoustic hearing provides information that is not transmitted by a conventional cochlear implant. In figure 5 we have plotted the performance of patients in EAS conditions for CNC words, consonants, vowels, sentences in quiet and sentences in noise (+10 and +5 dB SNR) against the performance, in the same test conditions, of 65 patients fit with a conventional cochlear implant whose CNC scores were 50% or better. In figure 6 we have plotted the performance of both groups of patients for melody recognition and voice discrimination. There is no evidence from any of the test conditions shown in figures 5 and 6 that EAS patients outperform the very best patients fit with a conventional cochlear implant. This is especially clear for sentences in noise, where performance is not constrained by a ceiling effect, e.g., sentences at +10 and +5 dB SNR, for CNC words and for within-gender voice discrimination. Thus, some patients who receive only E stimulation receive the same sum of information as patients who receive both E and A stimulation. This is a surprising outcome because it has been assumed, at least tacitly, that low-frequency hearing provides infor-

mation that is not coded, or not well coded, by a conventional cochlear implant.

It is the case, however, that the percentage of patients who receive the highest scores on any measure is much higher for EAS patients than for conventional cochlear implant patients. For example, 6% of conventional cochlear implant patients score 85% correct or better on sentences at +10 dB SNR. In contrast 33% of EAS patients achieve scores at this level. The difference in proportions is significant (Fisher's exact test, $p < 0.0003$). For melodies 11% of conventional patients achieve scores of 85% correct or better. Fifty-three percent of EAS patients achieve scores at this level. The difference in proportions is significant ($p < 0.0001$). Thus, while EAS does not push scores above those achieved by the very best conventional patients, the proportion of patients with very high scores is clearly greater for combined acoustic and electric stimulation.

Summary

Patients fit with a cochlear implant in one ear and a hearing aid in the other ear were presented with tests of speech and melody recognition and voice discrimination under conditions of electric stimulation, acoustic stimu-

lation and combined electric and acoustic stimulation. When acoustic information was added to electrically stimulated information, performance increased by 17–23 percentage points on tests of word and sentence recognition in quiet and sentence recognition in noise. On average, the EAS patients achieved higher scores on CNC words than patients fit with a unilateral cochlear implant. While the best EAS patients did not outperform the best patients fit with a unilateral cochlear implant, propor-

tionally more EAS patients achieved very high scores on tests of speech recognition than unilateral cochlear implant patients.

Acknowledgment

This work was supported by the following grants from the National Institute for Deafness and Other Communication Disorders: R01 DC 00654-16 to M.F.D. and DC 006538 to R.H.G.

References

- Brown CA, Bacon SP: The effect of fundamental frequency on speech intelligibility in simulated electric-acoustic listening. *J Acoust Soc Am* 2007;121:3039.
- Chang J, Bai J, Zeng F-G: Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. *IEEE Trans Biomed Eng* 2006;53:2598–2601.
- Childers D, Wu K: Gender recognition from speech. II. Fine analysis. *J Acoust Soc Am* 1991;90:1841–1856.
- Ching T, Incerti P, Hill M: Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear Hear* 2004;25:9–21.
- Clopper CG, Carter AK, Dillon CM, Hernandez LR, Pisoni DB, Clarke CM, Harnsberger JD, Herman R: The Indiana Speech Project: an overview of the development of a multi-talker multi-dialect speech corpus. Bloomington, Speech Research Laboratory, Indiana University, Research on Speech Perception Progress Report 2002;25:367–380.
- Dorman M, Dankowski K, McCandless G, Smith L: Identification of synthetic vowels by subjects using the Symbion multichannel cochlear implant. *Ear Hear* 1989;10:40–43.
- Gantz B, Turner C, Gfeller K, Lowder M: Preservation of hearing in cochlear implant surgery: advantages of combined electrical and acoustical speech processing. *Laryngoscope* 2005;115:796–802.
- Hartmann W, Johnson D: Stream segregation and peripheral channeling. *Music Percept* 1999;9:155–184.
- Helms J, Müller F, Schön F, et al: Evaluation of performance with the COMBI 40 cochlear implant in adults: a multicentric clinical study. *ORL* 1997;59:23–35.
- Kiefer J, Pok M, Adunka O, Sturzebecher E, Baumgartner W, Schmidt M, Tillein J, Ye Q, Gstöttner W: Combined electric and acoustic stimulation of the auditory system: results of a clinical study. *Audiol Neurotol* 2005;10:134–144.
- Kirk KI, Houston DM, Pisoni DB, Sprunger AB, Kim-Lee Y: Talker discrimination and spoken word recognition by adults with cochlear implants. Mid-Winter Meet Assoc Res Otolaryngol, St. Petersburg Beach, 2002.
- Kong Y-L, Stickney G, Zeng F-G: Speech and melody recognition in binaurally combined acoustic and electric hearing. *J Acoust Soc Am* 2005;117:1351–1361.
- Miller J: Auditory-perceptual interpretation of the vowel. *J Acoust Soc Am* 1989;85:2114–2134.
- Peters RW, Moore B: Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects. *J Acoust Soc Am* 1992;91:256–266.
- Peterson G, Lehiste I: Revised CNC lists for auditory tests. *J Speech Hear Disord* 1962;17:62–70.
- Qin M, Oxenham A: Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. *J Acoust Soc Am* 2006;119:2417–2426.
- Smith D, Walters T, Patterson R: Role of glottal-pulse rate, vocal-tract length and original talker upon judgments of speaker sex and age. *J Acoust Soc Am* 2007;121:3135.
- Spahr A, Dorman M: Performance of implant patients fit with the CII and Nucleus 3G devices. *Arch Otolaryngol Head Neck Surg* 2004;130:624–628.
- Spahr A, Dorman M, Loisel L: Performance of patients using different cochlear implant systems: effects of input dynamic range. *Ear Hear* 2007;28:260–275.
- Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA: Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing. *J Acoust Soc Am* 2004;115:1729–1735.
- Tyler R, Wood E, Fernandes M: Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners. *J Acoust Soc Am* 1983;74:1190–1199.
- von Ilberg C, Kiefer J, Tillein J, Pfenningdorff T, Hartmann R, Sturzebecker E, Klinke R: Electric-acoustic stimulation of the auditory system. *ORL* 1999;61:334–340.