Auditory Training in Patients With Unilateral Cochlear Implant and Contralateral Acoustic Stimulation

Ting Zhang,[1] Michael F. Dorman,[1] Qian-Jie Fu,[2] and Anthony J. Spahr[1]

Objectives: It was hypothesized that auditory training would allow bimodal patients to combine in a better manner the low-frequency acoustic information provided by a hearing aid with the electric information provided by a cochlear implant, thus maximizing the benefit of combining acoustic (A) and electric (E) stimulation (EAS).

Design: Performance in quiet or in the presence of a multitalker babble at +5 dB signal to noise ratio was evaluated in seven bimodal patients before and after auditory training. The performance measures comprised identification of vowels and consonants, consonant-nucleus-consonant words, sentences, voice gender, and emotion. Baseline performance was evaluated in the A-alone, E-alone, and combined EAS conditions once per week for 3 weeks. A phonetic-contrast training protocol was used to facilitate speech perceptual learning. Patients trained at home 1 hour a day, 5 days a week, for 4 weeks with both their cochlear implant and hearing aid devices on. Performance was remeasured after the 4 weeks of training and 1 month after training stopped.

Results: After training, there was significant improvement in vowel, consonant, and consonant-nucleus-consonant word identification in the E and EAS conditions. The magnitude of improvement in the E condition was equivalent to that in the EAS condition. The improved performance was largely retained 1 month after training stopped.

Conclusion: Auditory training, in the form administered in this study, can improve bimodal patients' overall speech understanding by improving E-alone performance.

(Ear & Hearing 2012;33;e70–e79)

INTRODUCTION

As the audiological candidacy criteria for cochlear implantation continues to evolve, an increasing number of unilateral cochlear implant (CI) users have substantial residual low-frequency hearing in the contralateral ear (Cohen 2004; Dorman & Gifford 2010). For these CI users, low-frequency information (which is not well transmitted by CIs) can be provided via hearing aids (HAs) in the nonimplanted ear. Speech performance both in quiet and in noise significantly improves when CI patients have access to both electric (E) and acoustic (A) stimulation (e.g., Tyler et al. 2002; Ching et al. 2004; Hamzavi et al. 2004; Kong et al. 2005; Mok et al. 2006; Gifford et al. 2007; Dorman et al. 2008; Zhang et al. 2010).

The benefit of combining E and A stimulation (EAS) across two ears in a bimodal listening condition varies significantly across patients. This variability is reflected not only in individual differences in EAS benefit but also in the time course of adaptation to the novel stimulation patterns provided by HAs and CIs. Some patients achieve substantial benefit, with sentence recognition in noise improving by 30 to 40 percentage points and monosyllablic word recognition in quiet improving by 20 to 30 percentage points, when acoustic stimulation is added to electric stimulation (e.g., Gstoettner et al. 2006, 2008; Gifford et al. 2007; Dorman et al. 2008; Helbig et al. 2008; Zhang et al. 2010). Other patients receive much less, or no, benefit even after extensive experience with EAS (e.g., Hamzavi et al. 2004; Kong et al. 2005). There have also been cases in which EAS produced poorer speech performance than E-alone stimulation (e.g., Tyler et al. 2002; Ching et al. 2004; Mok et al. 2006).

Besides the high variability in EAS benefit, individual patients also differ in terms of the time course of adaptation to EAS. After initial activation of the CI, EAS patients must adapt to differences between the patterns of activation produced by acoustic versus electric stimulation. Many studies have tracked performance over time in EAS patients. These longitudinal studies have shown that most EAS benefit occurs in the first 6 months of use (e.g., Shallop et al. 1992; Dooley et al. 1993; Gstoettner et al. 2006). However, continued improvement has been observed over 12 months of use for some EAS patients (e.g., Shallop et al. 1992; Gstoettner et al. 2008; Helbig et al. 2008).

One possible explanation for the variability in EAS benefit is variability in basic auditory function (e.g., audibility and frequency resolution) in the region of residual acoustic hearing (Turner et al. 2004; Kong et al. 2005). It is reasonable to suppose that ears with better hearing (i.e., better audibility and frequency resolution) would provide more EAS benefit than ears with poorer hearing (Gantz et al. 2009; Zhang et al. 2010). Although the basic auditory function in the region of residual acoustic hearing may be a contributing factor to the variability in EAS benefit, recent studies have demonstrated that the majority of EAS benefit is derived from low-frequency information from the voice fundamental frequency (F0) region containing voicing, amplitude envelope, and pitch-change cues (Kong & Carlyon, 2007; Brown et al. 2009; Zhang et al. 2010). If the addition of acoustic information from the F0 region alone is sufficient to improve speech performance with EAS relative to that with E-alone stimulation, and given that most EAS patients have substantial low-frequency residual hearing at and below 250 Hz, then why is there such a large variability in EAS benefit? EAS patients not only show a peripheral limitation in the processing of acoustic information, but may also differ in their ability to integrate the acoustic information from the F0 region and the electric information to achieve EAS benefit. During the initial period of combined use of an HA with a CI, EAS patients must adapt not only to the spectrally degraded and shifted speech patterns provided by CIs, but also to the perceptual dissimilarities between simultaneous acoustic and electric stimulation (Dooley et al. 1993; Ching et al. 2004). Although the long-term use of both devices may help EAS patients accommodate acoustically and electrically evoked speech patterns via "passive" learning, passive adaptation is not likely to be optimal. "Active" auditory training may improve patients' use of acoustic F0 cues and, in turn, maximize the benefit of EAS.

---

[1] Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona, USA; and [2] House Research Institute, Los Angeles, California, USA.

0196/0202/12/3306-0e70/0 • Ear & Hearing • Copyright © 2012 by Lippincott Williams & Wilkins • Printed in the U.S.A.

Auditory training, an important facet of aural rehabilitation, has been shown to be effective in improving speech-recognition skills of HA users by teaching hearing-impaired individuals to use in a better manner active listening strategies (e.g., hearing with attention and intention) to improve psychosocial function (e.g., Walden et al. 1981; Montgomery et al. 1984; Rubinstein & Boothroyd 1987; Sweetow & Palmer 2005). Recently, Fu et al. have reported encouraging results for the effectiveness of auditory training in improving CI users' auditory perception, including difficult listening tasks (e.g., speech in noise, telephone speech, music, etc.) (e.g., Fu et al. 2004, 2005; Fu & Galvin 2007, 2008; Oba et al. 2011). More importantly, the improved performance was largely retained after training had been stopped and the training benefit generalized to a variety of listening conditions. The training benefit for CI users can be affected by training protocol and materials (Fu et al. 2005; Li & Fu 2007; Stacey & Summerfield 2007, 2008; Loebach & Pisoni 2008; Stacey et al. 2010). In general, training (e.g., phoneme-recognition training) focusing on listeners' attention to acoustic contrasts/differences between stimuli targets "bottom-up" processes. Training that focuses on listeners' attention to lexically meaningful or contextual cues (e.g., keyword-in-sentence training) targets "top-down" processes. Both the bottom-up approach and the top-down approach have been shown to be effective for improving speech understanding (Fu et al. 2005; Li & Fu 2007; Stacey & Summerfield 2007, 2008; Fu & Galvin 2008; Loebach & Pisoni 2008; Stacey et al. 2010).

Because patients typically have more difficulty with phoneme identification than with word and sentence identification, we chose a bottom-up training approach rather than a top-down training approach to maximally challenge listeners to improve their auditory perception. The goal of the present study was to investigate the effect of a phoneme-recognition training on EAS benefit in patients with a CI in one ear and an HA in the other. To properly evaluate training outcomes, tasks that involve more bottom-up auditory processing, such as vowel, consonant, and consonant-nucleus-consonant (CNC) word identification, were included in baseline measures. It is likely that central pattern processing (i.e., top-down auditory processing) also plays a role in CNC word identification (e.g., the effect of lexical neighborhood density on word identification) (Krull et al. 2010). To document possible generalization of bottom-up segmental cue training to sentence recognition, which involves more top-down auditory processing, AzBio sentence identification was also included in the baseline measures. Last, to evaluate the benefit of adding low-frequency acoustic information to electric signal before and after auditory training, it was important to collect baseline measures that were sensitive to additional pitch-related cues provided by acoustic hearing. Therefore, voice gender and emotion identification were included in the baseline measures. The aims of this study were (1) to determine whether "active" auditory training would improve bimodal patients' ability to use the acoustic information in the F0 region and, in turn, maximize EAS benefit and (2) to determine whether phoneme-recognition training, in a closed-set task, would yield improvements in a variety of listening tasks including open-set tasks (e.g., sentence identification) and pitch-related tasks (e.g., voice gender and emotion identification).

MATERIALS AND METHODS

Subjects

Seven postlingually deafened CI adult users were recruited. All subjects had residual hearing in the nonimplanted ear. The individual audiogram for the nonimplanted ear is shown in Figure 1. The order of subject number was organized according to the amount of residual hearing (average pure-tone threshold for frequencies ≤ 1000 Hz) subjects had in the nonimplanted ear with S1 having the greatest and S7 having the least amount of residual hearing. Table 1 displays demographic information for each subject. At the time of testing, all subjects had at least 2 years of passive learning experience with their CI (mean = 4.3 years, SD = 3.4 years) and did not receive any active auditory training after cochlear implantation. All subjects gave informed consent to participate in this study and were compensated for their time. Informed consent procedures were approved by the Institutional Review Board at Arizona State University.

General Testing and Training Timeline

Because of the high variability in CI listener's performance, it is difficult to separate within-subject training effects from across-subject variability. Therefore, a "within-subject" control procedure (instead of an experimental control group) was adopted, in which each subject served as their own experimental control. Extensive baseline performance measures were obtained for the within-subject control procedure to independently estimate "procedural learning effects" (i.e., task familiarization) from "perceptual learning effects" (i.e., true active learning) for each subject. Baseline speech performance was repeatedly measured once per week for 3 weeks (pretraining). After the baseline measures, subjects trained at home on loaner laptops loaded with a custom training software (Sound Express; House Research Institute, Los Angeles, CA) for approximately 60 minutes per day,

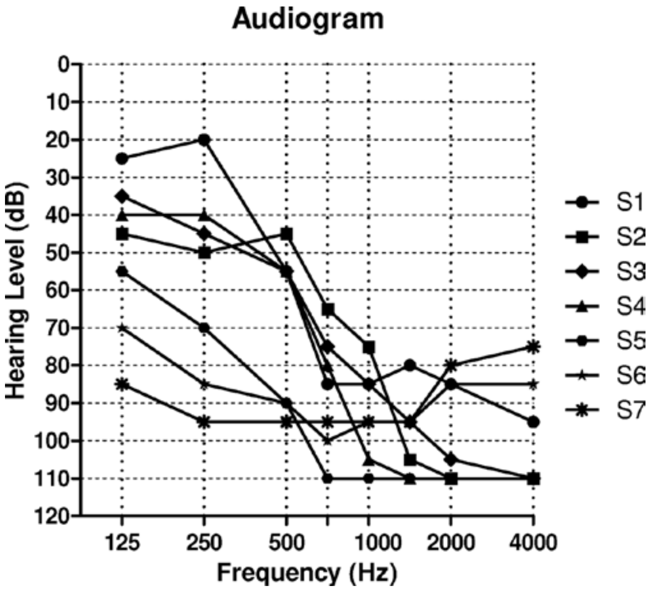

Fig. 1. Individual audiograms for the nonimplanted ear. The order of subject number was organized according to the residual hearing (average pure-tone threshold for frequencies ≤ 1000 Hz) subjects had in their nonimplanted ear with S1 having the greatest and S7 having the least amount of residual hearing.

TABLE 1. Subject demographics

Listener	Age (yrs)	Gender	Etiology	Duration of Hearing Loss* (NIE) (yrs)	Duration of Hearing Loss* (IE) (yrs)	CI Experience (yrs)	CI Device	HA Experience (NIE) (yrs)	HA Model	HA Usage (% Waking hr)
1	63	M	Unknown	23	23	2	Harmony	23	Phonak BTE	100
2	54	F	Unknown	46	46	3	Harmony	45	Phonak BTE	100
3	65	F	Unknown	16	16	7	Harmony	16	Senso BTE	100
4	66	M	Unknown	16	16	2	Harmony	13	Oticon BTE	100
5	65	F	Hereditary	58	58	10	CII BTE	48	Simens BTE	100
6	78	F	Measles	76	76	2	Harmony	44	Phonak BTE	100
7	51	F	Noise exposure	19	19	2	Harmony	14	Argosy BTE	100

NIE, nonimplanted ear; IE, implanted ear; CI, cochlear implant; HA, hearing aid.

**Duration of hearing loss was defined as duration of time since patients first noticed inability to understand a conversation on the telephone as being an indication of significant hearing loss.*

5 days per week, for 4 weeks. Performance was reevaluated after completing the 4th week of training (posttraining) and was remeasured 4 weeks after the training stopped (follow-up).

Test Methods and Materials

All baseline performance was measured in quiet (for one subject S6) or in the presence of a competing multiple-talker babble at +5 dB signal to noise ratio (SNR) (for the other six subjects). The reason for measuring speech performance in different listening backgrounds was to set the pretraining speech performance for each subject at a range of 20 to 80% to eliminate ceiling effects that might influence posttraining outcome measures. Speech stimuli were presented at 65 dB SPL in the free field via a single loudspeaker placed in front of the subject (0-degree azimuth) at a distance of 1 m. All baseline performance was evaluated in three listening conditions: (1) electric stimulation alone (E alone), (2) acoustic stimulation alone (A alone), and (3) combined electric and acoustic stimulation (EAS). Subjects were tested using their own HA and CI and their “everyday” programs. Before testing, subjects were presented with sample test items and they reported to the examiner that the sounds were “comfortably loud” and, in the case of EAS testing, that the acoustic and electric signals were balanced.

Baseline performance was evaluated using six sets of test materials, including four closed-set tasks and two open-set identification tasks. The four closed-set tasks included vowel and consonant identification, and voice gender and emotion identification. Vowel identification was measured in a 12-alternative paradigm. Vowel stimuli included 12 medial vowel tokens presented in h/V/d context produced by five male and five female talkers and drawn from the set recorded by Hillenbrand et al. (1995). Consonant identification was measured in a 20-alternative paradigm. Consonant stimuli included 20 medial consonant tokens presented in a/C/a format produced by five male and five female talkers and drawn from the set recorded by Shannon et al. (1999). Voice gender identification was measured in a two-alternative paradigm. The tokens were the same as those used in the vowel test. Emotion identification was measured in a five-alternative identification paradigm. Emotion stimuli included 40 semantically neutral sentences to convey five target emotions (angry, happy, sad, anxious, and neutral), produced by one male and one female talker (2 Talkers × 5 Emotions × 4 Sentences = 40 Sentences) and drawn from the database recorded at the House Research Institute.

The two open-set tasks included monosyllabic word and sentence identification. Word identification was tested using the CNC word lists (Peterson & Lehiste 1962). The materials included 10 phonemically balanced lists of 50 words recorded by a single male talker. Sentence identification was tested using the AzBio sentences organized into 33 lists of 20 sentences (Spahr et al. 2012). Sentences composed of 6 to 10 words were recorded by four talkers (two men and two women) using a casual speaking style. The sentence lists were constructed to have an equal number of sentences spoken by each of four speakers (two men and two women) and to have a consistent overall level of intelligibility. Both monosyllabic word and sentence understanding were evaluated in a total of 15 conditions (3 stimulation conditions [E, A, and EAS] × 5 performance measures [pretraining × 3, posttraining, and follow-up]). For both words and sentences, the list-to-condition assignments were randomized for each listener. However, with 10 CNC word lists and 15 conditions, 5 word lists used in the pretraining measures were assigned again to the posttraining and follow-up measures. There was a gap of at least 2 months for subjects to listen to the same word list and, therefore, the likelihood of the familiarization of the repeated word list was minimal. A novel sentence list was used in each condition. Before testing, listeners were allowed a brief practice session in each condition. The condition order was randomized and counterbalanced among listeners.

Training Materials and Methods

After the baseline measures were completed, training was conducted at home, using loaner laptops loaded with a custom training software (Sound Express). Subjects trained while listening to stimuli played back via computer speakers. Extensive training was provided to each subject on how to use the software, how to set up the computer speaker, and how to set the listening level. Subjects were also instructed to use the same CI and HA programs and volume control settings as those used for testing. Each subject’s training sessions were recorded in the Sound Express software, including performance and total training time. The software also allowed an experimenter to remotely access the training record to monitor each subject’s training frequency and performance. Subjects trained for approximately 60 minutes per day, 5 days per week, for 4 weeks, for a total of 20 training hours.

A phoneme-contrast protocol was used in auditory training, in which subjects specifically focused on the differences in acoustic features among vowels or consonants to facilitate bottom-up perceptual learning of speech phonemes. For vowels, acoustic speech features included first and second formant frequencies (F1 and F2) and duration; for consonants, speech features included voice, manner, and place of articulation (Miller & Nicely 1955). Monosyllabic words produced by different talkers (recorded at the House Research Institute) were the primary training materials used in the training exercises. These monosyllabic words (>1000 monosyllabic and 200 nonsense words) were different from those used in the test database. Six subjects trained to identify initial, medial, or final consonants and medial vowels in the presence of a speech babble noise. During the identification training, a monosyllabic word was played in the presence of a speech babble, and the subject responded by clicking on one of four choices shown onscreen (e.g., "Jane," "Joan," "John," and "June"). The response choices differed by only one phoneme (i.e., initial, medial, or final consonants, and medial vowels). The SNR was automatically adjusted according to a subject's responses. If the subject answered correctly, visual feedback was provided and the SNR was automatically increased by 2 dB. If the subject answered incorrectly, auditory and visual feedback were provided (allowing subjects to compare their response with the correct response) and the SNR was decreased by 2 dB. One subject (S6) was trained to identify vowels and consonants in quiet. The protocol was similar to the phoneme-in-noise protocol described earlier. This subject began consonant/vowel identification training in which the speech-feature differences among consonants and vowels were gradually reduced as the subject's performance improved (i.e., from "can" versus "fan" to "can" versus "pan"). As the subject's performance improved, the acoustic differences between response choices were reduced or the number of response choices was increased.

RESULTS

All subjects completed the specified testing and training. The total time spent on testing ranged from 11 to 13 hours with a mean of 12 hours, and the total time spent on training ranged from 962 to 1271 minutes with a mean of 1078 minutes (18 hours). To independently estimate the procedural learning effect, pretraining performance was measured once per week for 3 weeks. The Kruskal–Wallis one-way analysis of variance, with test session as a factor, was performed on the data from the pretraining measures within each stimulation condition and each test. There was no significant effects of test session for any of the stimulation conditions and test materials, suggesting no significant procedural learning effect observed in the pretraining performance measures. Therefore, the data were collapsed across the three pretraining measures to calculate the pretraining performance within each stimulation condition and test, which was then compared with the posttraining and follow-up performance for each subject. Friedman two-way analysis of variance was performed on the data for each test with two within-subject factors: training (pretraining, posttraining, and follow-up) and stimulation mode (E, A, and EAS). If the overall p value obtained from the Friedman test was less than 0.05, then post hoc pairwise comparisons were done by Wilcoxon signed rank test. Statistical details are provided in Table 2.

After training, speech-recognition performance (vowel and consonant identification, CNC word and AzBio sentence recognition) improved for six of seven subjects. There was, however, considerable variability in the magnitude of improvement. Pitch-related performance (voice gender and emotion identification) did not improve for any of the seven subjects. One subject (S5) showed no improvement on any test measure. This subject had the longest period of experience with a CI (10 years).

Figure 2 shows individual and mean pretraining, posttraining, and follow-up percentage correct scores as a function of stimulation condition for vowel identification. Posttraining performance significantly improved for the E (mean = 9.6 %; range = 2.3–14.3%) and EAS (mean = 8.6 %; range = 1.0–17.3%) conditions. Further analyses on the confusion matrices for vowel identification before and after training revealed that the perception of vowel place (F2) was significantly improved after the training both for the E (mean = 8.4%; range = −0.5 to 20.1%) and EAS conditions (mean = 11.9%; range = 0.5–17.5%). Figure 3 shows individual and mean pretraining, posttraining, and follow-up percentage correct scores as a function of stimulation condition for consonant identification. Posttraining performance significantly improved for the E (mean = 11.9 %; range = 3.3–20.7%) and EAS (mean = 9.8%; range = 2.3–14.0%) conditions. Further analyses on the confusion matrices for consonant identification before and after training revealed that the perception of consonant voicing and manner of articulation was significantly improved after the training both for the E (voicing: mean = 16.5%, range = 0.5–26.5%; manner: mean = 10.4%, range = 2.5–25%) and EAS conditions (voicing: mean = 13.4%, range = 0.5–34.5%; manner: mean = 10.1%, range = −4.5 to 26.5%). Figure 4 shows individual and mean pretraining, posttraining, and follow-up percentage correct scores as a function of stimulation condition for CNC word identification. Posttraining performance significantly improved for the E (mean = 13.6%; range = 2.7–33%) and EAS (mean = 14.9%; range = 1.3–33.3%) conditions. For all aforementioned three tests, the magnitude of improvement in the EAS condition was equivalent to that in the E condition (mean differences between EAS score and E-alone score in pretraining, posttraining, and follow-up measures; $p > 0.05$; see Table 2), and the improvement was largely retained 1 month after training stopped (mean differences between posttraining and follow-up performance in the E and EAS conditions; $p > 0.05$; see Table 2).

Figure 5 shows mean pretraining, posttraining, and follow-up percentage correct scores as a function of stimulation condition for AzBio sentence identification. Posttraining performance improved for the E (mean = 6.7%; range = −7.3 to 17.6%) and EAS (mean = 8.3%; range = −4 to 21.2%) conditions. However, the improvement for AzBio sentence identification failed to reach statistical significance. Figure 5 shows mean pretraining, posttraining, and follow-up percentage correct scores as a function of stimulation condition for voice gender and emotion identification; there was no significant training benefit. There was also no significant training benefit observed for the A condition for any test material.

DISCUSSION

On average, the bottom-up segmental cue training resulted in a 10% improvement in vowel and consonant identification and CNC word identification in quiet or in noise for bimodal patients who had already extensive experience (at least 2 years) with their CI. The magnitude of the improvement in the E condition was equivalent to that

TABLE 2. Results of Friedman two-way analysis of variance and post hoc pairwise comparison (Wilcoxon signed rank test) as a function of test materials

	Friedman Two-Way Analysis of Variance* (Significant at 0.05)						Post Hoc Pairwise Comparison (Wilcoxon Signed Rank test)*							
							Training				Stimulation Mode			
	Training			Stimulation Mode				Mean Difference				Mean Difference		
Test Materials	A	E	EAS	Pretraining	Posttraining	Follow-Up	Pairwise	A	E	EAS	Pairwise	Pretraining	Posttraining	Follow-Up
Vowel	>0.05	**<0.01***	**<0.05***	**<0.05***	**<0.05***	**<0.05***	Post–Pre	3.2	**9.6***	**8.6***	E–A	**28.7***	**34.4***	**35.7***
							F–Pre	3.0	**10.2***	**12.2***	EAS–A	**32.1***	**36.7***	**41.1***
							F–Post	−0.7	0.6	3.7	EAS–E	3.4	2.3	5.4
Consonant	>0.05	**<0.05***	**<0.05***	**<0.05***	**<0.05***	**<0.05***	Post–Pre	2.0	**11.9***	**9.8***	E–A	**26.1***	**36.1***	**33.6***
							F–Pre	1.6	**9.0***	**10.3***	EAS–A	**32.1***	**40.0***	**40.9***
							F–Post	−0.4	−2.9	0.6	EAS–E	6	3.9	7.3
CNC word	>0.05	**<0.05***	**<0.05***	**<0.05***	**<0.05***	**<0.05***	Post–Pre	2.4	**13.6***	**14.9***	E–A	**28.5***	**39.7***	**38.0***
							F–Pre	2.1	**11.6***	**14.0***	EAS–A	**33.5***	**46.0***	**45.4***
							F–Post	−0.3	−2.0	−0.9	EAS–E	5.0	6.3	7.4
AzBio sentence	>0.05	>0.05	>0.05	**<0.05***	**<0.05***	**<0.05***					E–A	**30.7***	**35.9***	**33.9***
											EAS–A	**42.0***	**48.7***	**44.3***
											EAS–E	11.3	**12.8***	10.4
Voice gender	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05								
Emotion	>0.05	>0.05	>0.05	**<0.05***	**<0.05***	**<0.05***					E–A	3.4	1.9	3.9
											EAS–A	**11.3***	**10***	**12.4***
											EAS–E	**8.2***	8.1	**12.4***

The analysis was performed with two within-subject factors: training (pretraining, posttraining and follow-up) and stimulation condition (E, A, and EAS). The test materials included vowel/consonant identification, CNC word, and AzBio sentence identification, and voice gender and emotion identification. Significant main effects and post hoc comparisons are indicated in boldface and with asterisks.

**Significant at 0.05.*

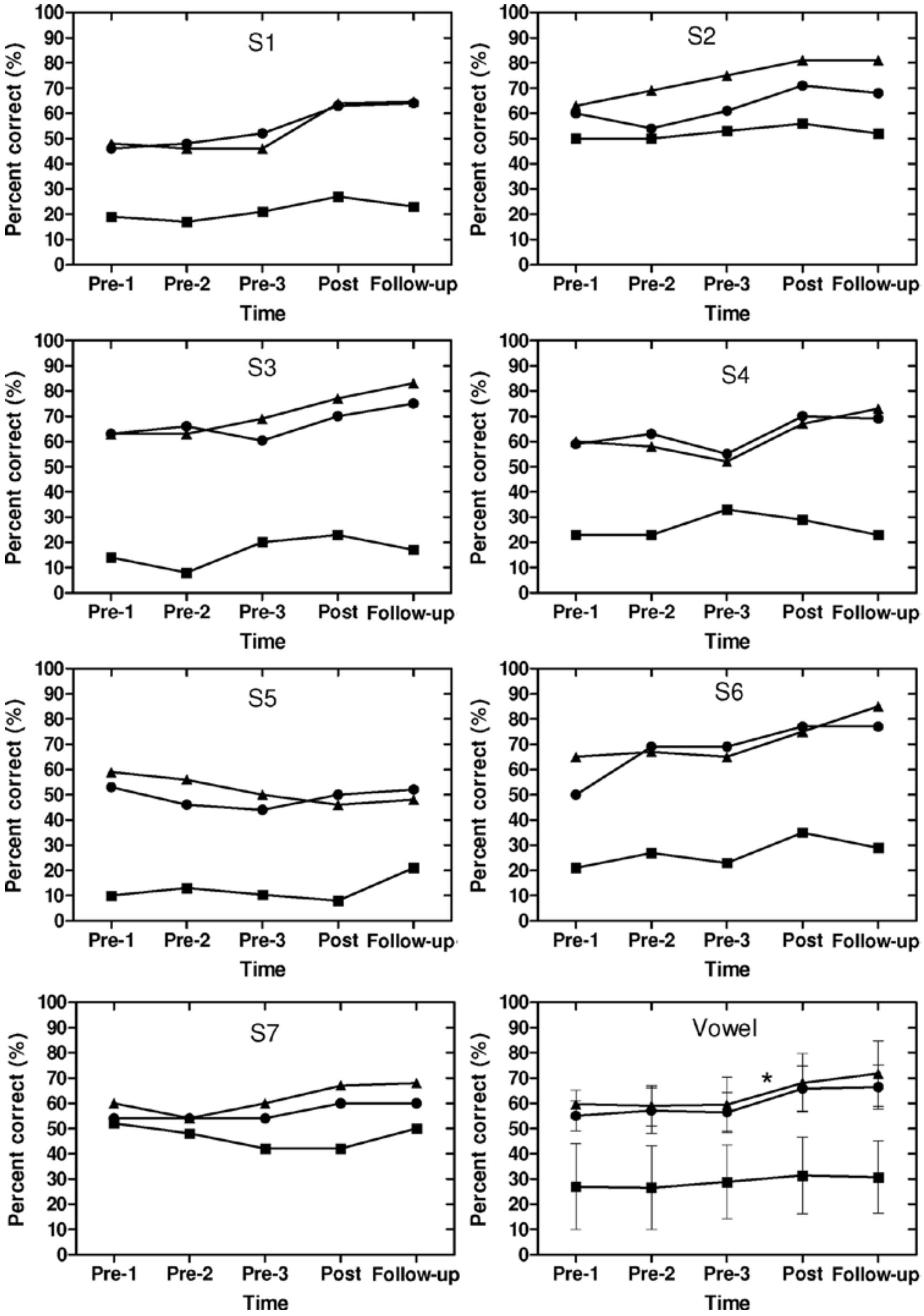

Fig. 2. Individual and mean performance as a function of stimulus condition (A, E, and EAS) for vowel identification in quiet (S6) or in the presence of a multitalker babble at +5 dB signal to noise ratio (other six subjects). Error bars indicate ± 1 SD. Asterisks (*) indicate statistical significance.

in the EAS condition, and the improvement was largely retained 1 month after training stopped. There was no significant training benefit observed for two pitch-related performance tasks—voice gender and emotion identification.

Although mean speech-recognition performance significantly improved with training, there was a large variability in the magnitude of improvement across subjects. Some subjects achieved a larger training benefit (e.g., 20–30% for CNC word recognition for S1 and S6), and others benefited modestly from the training (e.g., 8–12% for CNC word recognition for S2, S3, S4, and S7). Subject 5 did not demonstrate any training benefit in any speech-recognition task. Nevertheless, most subjects' speech-recognition performance improved after training. All subjects had at least 2 years of experience with their CI and HA and, therefore, had a long passive learning experience with their devices and had enough time to adapt to perceptual dissimilarities between simultaneous acoustic and electric stimulation (e.g., Shallop et al. 1992; Gantz & Turner 2004; Gstoettner et al. 2008; Helbig et al. 2008). Subject 5, the only subject without training benefit, had the longest period of experience

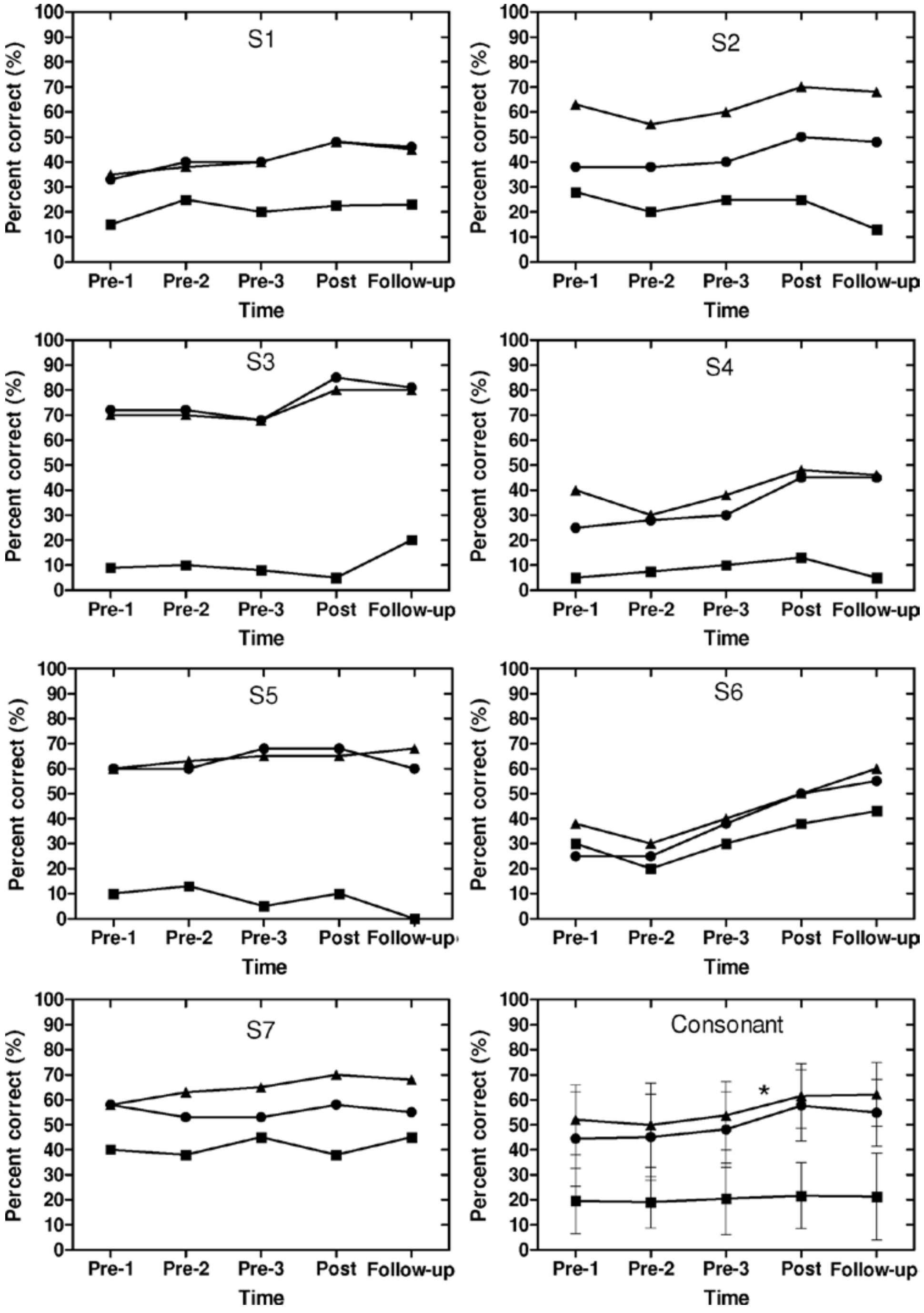

Fig. 3. Individual and mean performance as a function of stimulus condition (A, E, and EAS) for consonant identification in quiet (S6) or in the presence of a multitalker babble at +5 dB signal to noise ratio (other six subjects). Error bars indicate ± 1 SD. Asterisks (*) indicate statistical significance.

(10 years) with her CI and HA, and she was the most active and self-initiated listener among the seven subjects, who had been frequently engaging in learning experience (e.g., by audio-book listening) after cochlear implantation. Given that the improvement in speech perception mostly occurred within the 1st year postimplantation and might continue up to 5 years postimplantation (Spivak & Waltzman 1990; Loeb & Kessler 1995; Tyler et al. 1997), it was possible that S5's speech performance had reached a plateau after years of passive learning experience and, therefore, auditory training could not further improve her performance. Overall, however, our results were consistent with the previous studies that reported the effectiveness of auditory training in improving CI users' auditory perception (Fu et al. 2004, 2005; Fu & Galvin, 2007, 2008; Oba et al. 2011).

In addition to the variability in training benefit across subjects, there was a large variability in the magnitude of improvement across speech-recognition tasks. The magnitude of improvement was greatest for CNC word identification (mean = 14%), modest for vowel (mean = 9%) and consonant identification (mean = 10%), and the least for AzBio sentence

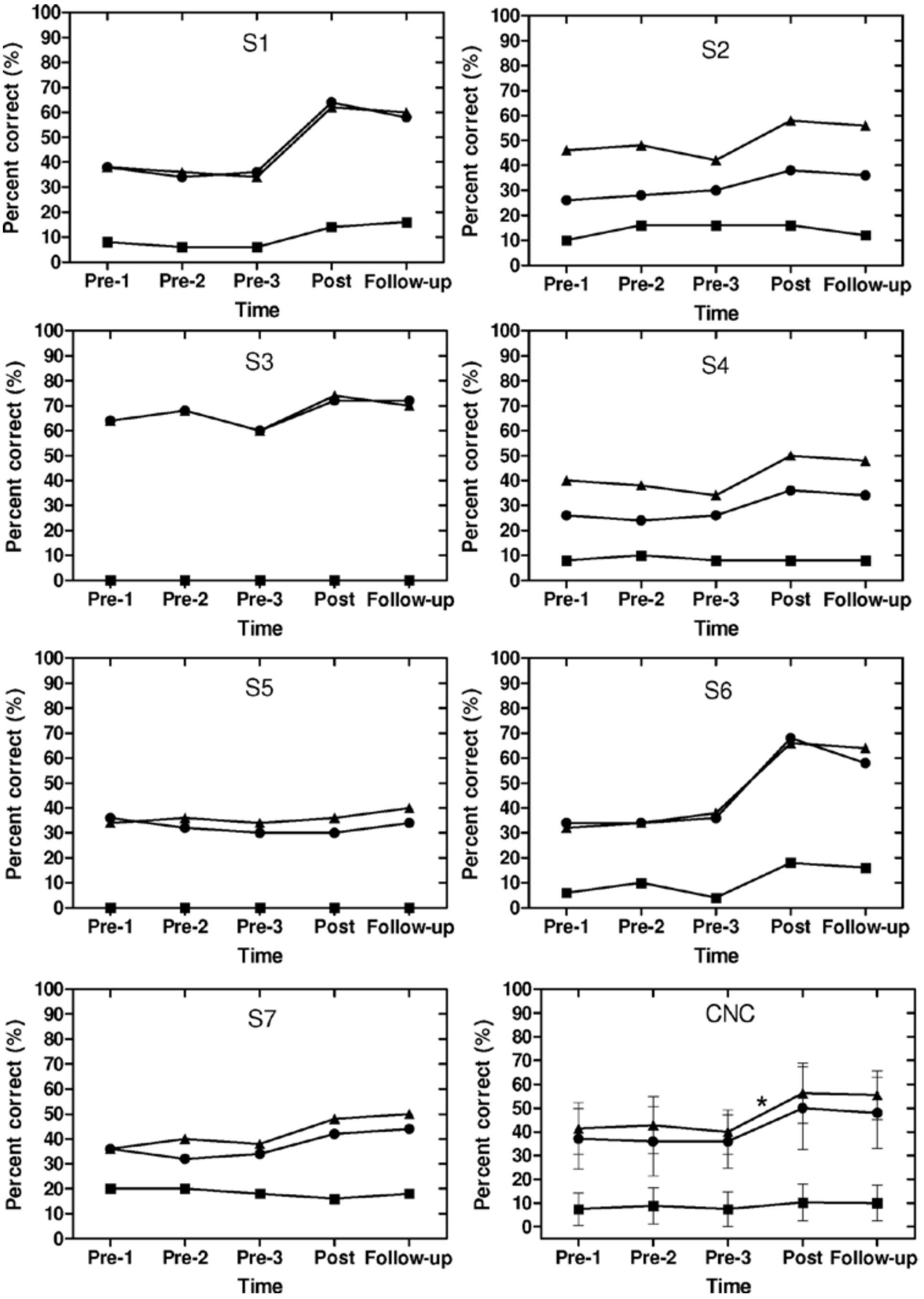

Fig. 4. Individual and mean performance as a function of stimulus condition (A, E, and EAS) for CNC word identification in quiet (S6) or in the presence of a multitalker babble at +5 dB signal to noise ratio(other six subjects). Error bars indicate ± 1 SD. Asterisks (*) indicate statistical significance.

identification (mean = 7%). The present auditory training used a bottom-up approach, which targeted on enhancing listeners' attention to phonemically relevant features to facilitate bottom-up learning in discriminating phonemic contrasts/differences among phonemes. Therefore, the phoneme-recognition training significantly improved vowel and consonant identification. Further analyses of the confusion matrices for vowel and consonant identification before and after training revealed that the perception of vowel place (F2) and consonant voicing and manner of articulation were significantly improved after the training both for the E and EAS conditions. In addition, improved perception of vowel place and consonant manner and voicing may have facilitated the top-down linguistic process of narrowing potential word candidates in a lexicon (e.g., Zue, 1985). Therefore, training also significantly improved CNC word identification. However, training minimally enhanced the higher-level linguistic processing at the level of connected speech (AzBio sentence recognition). Fu and Galvin (2008) reported that a top-down approach (targeting contextual cues available with sentence training) provided a greater training-in-noise benefit than a bottom-up approach. Therefore, it is possible that AzBio sentence recognition might have been

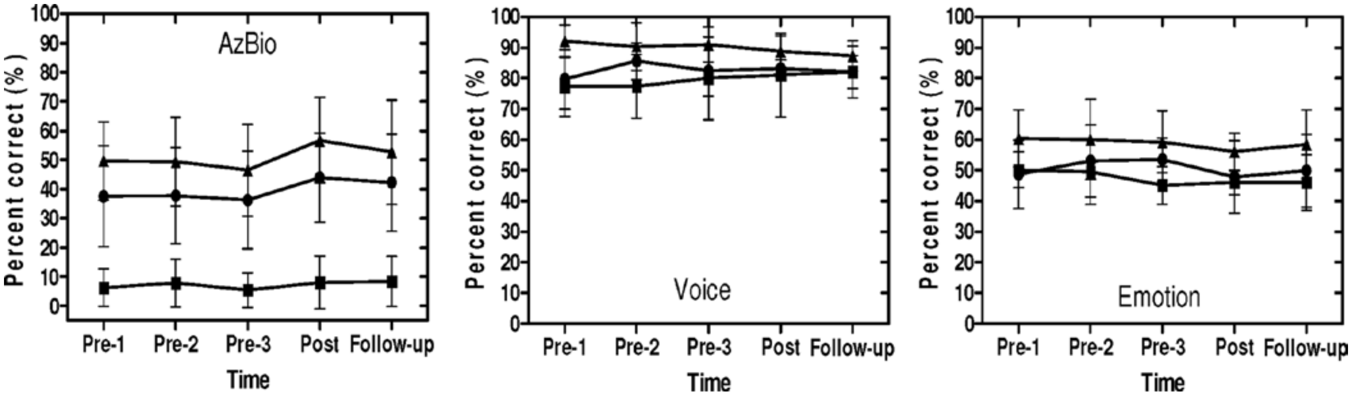

Fig. 5. Mean performance as a function of stimulus condition (A, E, and EAS) for sentence identification, voice gender, and emotion identification in quiet (S6) or in the presence of a multitalker babble at +5 dB signal to noise ratio(other six subjects). Error bars indicate ± 1 SD.

significantly improved if the top-down training approach had been used. For voice gender and emotion identification, it is not surprising that training focused on phonetic contrasts among vowels and consonants did not improve performance.

The magnitude of the training benefit in the E condition was equivalent to that in the EAS condition, and there was no significant training benefit observed in the A condition for any of the test materials. The results suggest that the training benefit is likely a result of auditory learning for the electric signal and that the EAS benefit cannot be further maximized by the auditory training. Given that the majority of the EAS benefit is derived from low-frequency information from the F0 region (Brown et al. 2009; Zhang et al. 2010) and most subjects recruited for the present study had substantial low-frequency residual hearing in this region, one possible explanation for the absence of maximizing EAS benefit with auditory training is the limitation of processing the acoustic information at suprathreshold level (i.e., frequency selectivity in the region of residual acoustic hearing). Zhang et al. (2010) reported that individuals achieving a larger amount of EAS benefit across two ears were able to process better the acoustic information in the frequency domain. Subjects in the present study had limited benefit of combining EAS across two ears, so the frequency selectivity in the region of acoustic hearing in this group of subjects was likely to be poor. Therefore, the training effect on EAS benefit might have been compromised by the poor frequency selectivity in the region of acoustic hearing. It is reasonable to hypothesize that maximizing EAS benefit with auditory training would have been observed if a group of subjects achieving a larger amount of EAS benefit (i.e., with better frequency selectivity in the region of acoustic hearing) had been recruited for the present study. However, this does not seem to be the case. Subject 2 achieved substantial benefit with consonant, word, and sentence recognition in noise improving by 20 to 30 percentage points when acoustic stimulation was added to electric stimulation. However S2's EAS benefit was not maximized after auditory training. More research needs to be done to investigate whether the peripheral limitation at suprathreshold level (i.e., frequency selectivity in the region of acoustic hearing) may play a role in the auditory training effect in bimodal patients.

Another possible explanation for the absence of maximizing EAS benefit with auditory training is the time frame of training provided to the subjects. All subjects had extensive experience (at least 2 years) with their CI and HA before training started. As mentioned in the Introduction, most EAS benefit occurs in the first 6 months of use after cochlear implantation (e.g., Shallop et al. 1992; Dooley et al. 1993; Gstoettner et al. 2006). It is possible that training initiated at earlier stage after cochlear implantation would have been able to better facilitate the perception of signals elicited by acoustic and electric stimulation across two ears and, in turn, maximize the EAS benefit. Given the training outcomes from the present study, auditory training should be recommended for all bimodal patients after cochlear implantation, which can improve their overall speech understanding.

RESULTS

This study demonstrated that, on average, a phonemic-based auditory training resulted in a 10% improvement in vowel, consonant, and CNC word identification performance in the E and EAS conditions for bimodal patients who had extensive experience (at least 2 years) with their CI and HA. However, the auditory training in a closed-set task did not yield significant improvement in an open-set listening task (sentence identification) and two pitch-related listening tasks (voice gender and emotion identification). The magnitude of training benefit in the E condition was equivalent to that in the EAS condition, suggesting that the benefit was because of the central auditory learning effect for the electric signal but not the acoustic signal. The training benefit remained 1 month after training stopped. Although EAS benefit was not maximized by the auditory training, training in the form administered in this study can improve overall speech understanding of bimodal patients by improving their E-alone performance.

ACKNOWLEDGMENTS

This research was supported by awards from the National Institute on Deafness and Other Communication Disorders, to Ting Zhang (1 F32 DC010937-01A1) and to Michael Dorman (R01 DC DC010494-01).

Address for correspondence: Ting Zhang, PhD, 658 N. Poplar Court, Chandler, AZ 85226, USA. E-mail: ting.zhang@asu.edu

Received September 23, 2011; accepted April 4, 2012.

REFERENCES

Brown, C. A., Bacon, S. P., Scherrer, N. M. (2009). The effect of fundamental frequency on speech intelligibility in simulated electric-acoustic listening. *J Acoust Soc Am*, *121*, 3039.

Ching, T. Y., Incerti, P., Hill, M. (2004). Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear Hear*, *25*, 9–21.

Cohen, N. L. (2004). Cochlear implant candidacy and surgical considerations. *Audiol Neurootol*, *9*, 197–202.

Dooley, G. J., Blamey, P. J., Seligman, P. M., et al. (1993). Combined electrical and acoustical stimulation using a bimodal prosthesis. *Arch Otolaryngol Head Neck Surg*, *119*, 55–60.

Dorman, M. F., & Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *Int J Audiol*, *49*, 912–919.

Dorman, M. F., Gifford, R. H., Spahr, A. J., et al. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiol Neurootol*, *13*, 105–112.

Fu, Q. J., & Galvin, J. J., III. (2007). Perceptual learning and auditory training in cochlear implant recipients. *Trends Amplif*, *11*, 193–205.

Fu, Q. J., & Galvin, J. J., III. (2008). Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hear Res*, *242*, 198–208.

Fu, Q. J., Galvin, J. J., III, Wang, X., et al. (2005). Moderate auditory training can improve speech performance of adult cochlear implant users. *J Acoust Soc Am, 6* (ARLO), 106–111.

Fu, Q. J., Hsu, C. J., Horng, M. J. (2004). Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant patients. *Ear Hear, 25,* 501–508.

Gantz B. J., Hansen M. R., Turner C. W., et al. (2009) Hybrid 10 clinical trial: preliminary results. *Audiol Neurootol. 14* (Suppl 1), 32–38.

Gantz, B. J., & Turner, C. (2004). Combining acoustic and electric speech processing: Iowa/nucleus hybrid implant. *Acta Otolaryngol, 124*, 344–347.

Gifford, R. H., Dorman, M. F., McKarns, S., et al. (2007). Combined electric and contralateral acoustic hearing: Word and sentence intelligibility with bimodal hearing. *J Speech Lang Hear Res, 50*, 835–843.

Gstoettner, W. K., Helbig, S., Maier, N., et al. (2006). Ipsilateral electric acoustic stimulation of the auditory system: results of long-term hearing preservation. *Audiol Neurootol*, *11* (Suppl 1), 49–56.

Gstoettner, W. K., van de Heyning, P., O'Connor, A. F., et al. (2008). Electric acoustic stimulation of the auditory system: results of a multi-centre investigation. *Acta Otolaryngol*, *128*, 968–975.

Hamzavi, J., Pok, S. M., Gstoettner, W., et al. (2004). Speech perception with a cochlear implant used in conjunction with a hearing aid in the opposite ear. *Int J Audiol*, *43*, 61–65.

Helbig, S., Baumann, U., Helbig, M., et al. (2008). A new combined speech processor for electric and acoustic stimulation–eight months experience. *ORL J Otorhinolaryngol Relat Spec*, *70*, 359–365.

Hillenbrand, J., Getty, L. A., Clark, M. J., et al. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am*, *97,* 3099–3111.

Kong, Y. Y., & Carlyon, R. P. (2007). Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *J Acoust Soc Am*, *121*, 3717–3727.

Kong, Y. Y., Stickney, G. S., Zeng, F. G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *J Acoust Soc Am*, *117*(3 Pt 1), 1351–1361.

Krull, V., Chio, S., Kirk, K. I., et al. (2010). Lexical effects on spoken word recognition in children with normal hearing. *Ear Hear*, *31*, 102–114.

Li, T., & Fu, Q. J. (2007). Perceptual adaptation to spectrally shifted vowels: training with nonlexical labels. *J Assoc Res Otolaryngol, 8*, 32–41.

Loeb, G. E., & Kessler, D. K. (1995). Speech recognition performance over time with the Clarion cochlear prosthesis. *Ann Otol Rhinol Laryngol Suppl*, *166*, 290–292.

Loebach J. L., & Pisoni, D. B. (2008). Perceptual learning of spectrally degraded speech and environmental sounds. *J Acoust Soc Am, 123*, 1126–1139.

Miller, G. & Nicely, P. (1995). An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am, 27*, 338–352.

Mok, M., Grayden, D., Dowell, R. C., et al. (2006). Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. *J Speech Lang Hear Res*, *49*, 338–351.

Montgomery, A., Walden, B., Schwartz, D., et al. (1984). Training auditory-visual speech recognition in adults with moderate sensorineural hearing loss. *Ear Hear*, *5*, 30–36.

Oba, S. I., Fu, Q. J., Galvin, J. J., III. (2011). Digit training in noise can improve cochlear implant users' speech understanding in noise. *Ear Hear*, *32*, 573–581.

Peterson, G. E., & Lehiste, I. (1962). Revised CNC lists for auditory test. *J Speech Hear Disorders, 27*, 62–70.

Rubinstein, A., & Boothroyd, A. (1987). Effect of two approaches to auditory training on speech recognition by hearing impaired adults. *J Speech Hear Res*, *30*, 153–160.

Shallop, J. K., Arndt, P. L., Turnacliff, K. A. (1992). Expanded indications for cochlear implantation: Perceptual results in seven adults with residual hearing. *J Spoken Lang Path Audiol*, *16*, 141–148.

Shannon, R. V., Jensvold, A., Padilla, M., et al. (1999). Consonant recordings for speech testing. *J Acoust Soc Am*, *106,* L71–L74.

Spahr, A. J., Dorman, M. F., Litvak, L. M., et al. (2012). Development and validation of the AzBio sentence lists. *Ear Hear*, *33*, 112–117.

Spivak, L. G., & Waltzman, S. B. (1990). Performance of cochlear implant patients as a function of time. *J Speech Hear Res*, *33*, 511–519.

Stacey, P. C., Raine, C. H., O'Donoghue, G. M., et al. (2010). Effectiveness of computer-based auditory training for adult users of cochlear implants. *Int J Audiol*, *49*, 347–356.

Stacey, P. C., & Summerfield, A. Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *J Acoust Soc Am*, *121*(5 Pt1), 2923–2935.

Stacey, P. C., & Summerfield, A. Q. (2008). Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally distorted speech. *J Speech Lang Hear Res*, *51*, 526–538.

Sweetow, R., & Palmer, C. V. (2005). Efficacy of individual auditory training in adults: A systematic review of the evidence. *J Am Acad Audiol*, *16*, 494–504.

Turner, C. W., Gantz, B. J., Vidal, C., et al. (2004). Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *J Acoust Soc Am*, *115*, 1729–1735.

Tyler, R. S., Parkinson, A. J., Wilson, B. S., et al. (2002). Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization. *Ear Hear*, *23*, 98–105.

Tyler, R. S., Parkinson, A. J., Woodworth, G. G., et al. (1997). Performance over time of adult patients using the Ineraid or nucleus cochlear implant. *J Acoust Soc Am*, *102*, 508–522.

Walden, B. E., Erdman, S. A., Montgomery, A. A., et al. (1981). Some effects of training on speech recognition by hearing-impaired adults. *J Speech Hear Res*, *24*, 207–216.

Zhang, T., Dorman, M. F., Spahr, A. J. (2010). Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear Hear*, *31*, 63–69.

Zue, V. (1985). The use of speech knowledge in automatic speech recognition. *Proc IEEE, 73*, 1602–1615.