

Effects of Computer-Assisted Speech Training on Mandarin-Speaking Hearing-Impaired Children

Jiunn-Liang Wu^a Hui-Mei Yang^a Yi-Hui Lin^a Qian-Jie Fu^b

^aDepartment of Otolaryngology, College of Medicine, National Cheng Kung University, Tainan, Taiwan;

^bDepartment of Auditory Implants and Perception, House Ear Institute, Los Angeles, Calif., USA

Key Words

Children · Cochlear implants · Hearing aids · Speech recognition

Abstract

The present study investigated whether moderate amounts of computer-assisted speech training can improve the speech recognition performance of hearing-impaired children. Ten Mandarin-speaking children (3 hearing aid users and 7 cochlear implant users) participated in the study. Training was conducted at home using a personal computer for half an hour per day, 5 days per week, for a period of 10 weeks. Results showed significant improvements in subjects' vowel, consonant, and tone recognition performance after training. The improved performance was largely retained for 2 months after training was completed. These results suggest that moderate amounts of auditory training, using a computer-based auditory rehabilitation tool with minimal supervision, can be effective in improving the speech performance of hearing-impaired children.

Copyright © 2007 S. Karger AG, Basel

Introduction

The hearing aid (HA) and cochlear implant (CI) provide hearing sensation to patients with severe or profound hearing loss. With advances in HA and CI technol-

ogy, the overall speech recognition performance of hearing-impaired patients has steadily improved [Zeng, 2004]; many patients receive great benefit, and are even capable of telephone conversations with friends and family. However, considerable variability remains in individual patient outcomes. This variability in patient outcomes is reflected not only in individual differences in speech performance, but also in the time course of adaptation to the novel speech patterns provided by the HA or CI. While some patients may easily and quickly adapt to these new speech patterns, others may require an extensive learning period. It is generally accepted that electronic devices alone will not fully meet the needs of hearing-impaired patients, and that auditory rehabilitation would enhance the benefits provided by HA and CI devices.

Auditory training, an important facet of aural rehabilitation, has been shown to be effective in the rehabilitation of children with central auditory processing disorders [Hesse et al., 2001], children with language-learning impairment [Merzenich et al., 1996], and HA users [Sweetow and Palmer, 2005]. Several studies have also assessed the effects of auditory training on speech recognition by poor-performing CI patients [Busby et al., 1991; Dawson and Clark, 1997; Fu et al., 2005a]. Busby et al. [1991] examined the effects of 10 1-hour speech perception training sessions (1–2 sessions per week); 3 prelingually deafened CI users (2 adolescents and 1 adult) participated in that experiment. Only minimal changes in speech performance were observed after training was completed;

the subject with the greatest improvement was implanted at an earlier age, and therefore had a shorter period of deafness. Dawson and Clark [1997] also reported the data for vowel recognition training in 5 CI users. Each subject had been deaf for at least 4 years prior to implantation, and none had achieved open-set speech recognition. Training consisted of 1 50-min training session per week for 10 weeks. Following training, 4 of the 5 subjects showed improvement in some measures; this improvement was retained on subsequent testing 3 weeks after training was completed.

Recently, Fu et al. [2005a] have reported more encouraging results for auditory training in 10 adult CI patients. The study was proposed to verify two primary hypotheses regarding speech perception training with CI patients: (1) CI users' poor speech recognition performance can be improved by moderate auditory training and (2) auditory rehabilitation can be performed with minimal supervision using a computer program installed on patients' home computers. The training program targeted simple vowel and consonant contrasts. Subjects trained at home for 1 h per day, 5 days per week, for a period of 1 month or longer, and returned to the lab regularly for re-testing of vowel, consonant and sentence recognition. The results showed significant improvement in all patients' speech recognition performance. On average, vowel recognition scores improved from 24 to 40% correct, and consonant recognition scores improved from 25 to 39% correct, after training was completed. Although patients' sentence recognition was not directly trained, sentence recognition also improved from 28 to 56% correct.

The results from Fu et al. [2005a] suggest that moderate amounts of auditory training, conducted at home using speech-training software, were effective in improving CI patients' speech understanding. By providing an auditory-only training environment, as well as guidance and feedback, the computer-based rehabilitation tool helped CI patients to better accommodate to the speech patterns delivered by the implant device. While previous training studies have evaluated the benefits of computer-assisted speech training for HA users and adult CI users, the utility of computer-assisted speech training has yet to be systematically evaluated in hearing-impaired children who use HA and/or CI devices. For hearing-impaired children, auditory rehabilitation is critical to hearing and speech development. However, access to auditory rehabilitation is limited, as rehabilitation programs are mainly provided by hospitals or hearing clinics, requiring significant amounts of time and expense. Computer-assist-

ed speech training may augment traditional 'hands-on' rehabilitation approaches by providing greater flexibility with minimal costs and supervision. The goal of the present study is to examine whether home training using auditory training software can improve the speech recognition performance of hearing-impaired children.

Most previous speech training studies have focused on improving recognition of sentences or segmental information. Relatively few studies have focused on improving recognition of suprasegmental information. In tonal languages such as Mandarin Chinese, the tonality of a syllable is lexically important [Lin, 1988]. Previous studies have shown that Chinese sentence recognition is highly correlated with tone recognition [Fu et al., 1998, 2004]. The fundamental frequency (F_0) contour contributes most strongly to tone recognition [Lin, 1988]. However, other temporal cues that covary with tonal patterns (e.g. vowel duration, amplitude contour and periodicity fluctuations) also contribute to tone recognition, especially when F_0 cues are reduced or unavailable [Fu et al., 1998; Fu and Zeng, 2000]. Because F_0 information is not explicitly coded in contemporary CI devices, CI users have to rely on spectral and temporal envelope cues for tone recognition. Temporal envelope cues have been shown to provide moderate levels of tone recognition in CI users [Fu et al., 2004; Hsu et al., 2000]. While auditory training can significantly improve CI users' phoneme recognition (i.e., segmental speech cues), it is unclear whether similar training can improve Chinese tone recognition (i.e., suprasegmental speech cues). Another goal of the present study is to investigate whether Chinese tone recognition can be improved by moderate auditory training in hearing-impaired children.

Methods

Subjects

Ten congenitally deafened children (3 HA users and 7 CI users) participated in the present study. The inclusion criteria for all participants were: (1) subjects had to be at least 5 years old, and (2) subjects had to have at least 2 months of experience with their device. All subjects were native speakers of Mandarin Chinese. Table 1 contains relevant information on the 10 subjects.

Testing Materials and Procedure

The speech materials for vowel, consonant, and Chinese tone recognition tests were recorded at the House Ear Institute. All speech stimuli were sampled at a 22-kHz sampling rate, without high-frequency preemphasis. One male and 1 female speaker each produced 4 tones (tone 1: flat, tone 2: rising, tone 3: falling-rising, tone 4: falling) for 6 Mandarin Chinese single-vowel syllables (/a/, /o/, /e/, /i/, /u/, /ü/), resulting in a stimulus set of 48 speech tokens.

Table 1. Subject demographics for all research participants

Subject	Age years	Gender	Type	Device	Strategy	Use of device years
S1	8.98	M	CI	Nucleus-24M	ACE	1.72
S2	6.91	F	CI	Clarion C-II	HiRes	0.34
S3	7.74	F	CI	Nucleus-24M	ACE	0.16
S4	10.88	M	CI	Nucleus-24M	ACE	3.64
S5	8.72	F	CI	Clarion C-II	HiRes	1.23
S6	8.14	F	CI	Nucleus-24M	ACE	6.34
S7	7.95	M	CI	Nucleus-24M	ACE	5.23
S8	7.69	F	HA	Oticon superpower II		5.75
S9	7.49	M	HA	Oticon superpower II		4.56
S10	5.87	M	HA	Oticon superpower II		4.89

ACE = Advanced Combination Encoders; HiRes = high resolution.

Vowel recognition was measured using a 6-alternative identification paradigm. Chinese tone recognition was measured with the vowel stimulus set, using a 4-alternative identification paradigm. For consonant test materials, 1 male and 1 female speaker each produced tone 1 for 19 Mandarin Chinese syllables (/a, ba, pa, ma, fa, da, ta, la, ga, ka, ha, za, ca, sa, zha, cha, sha, ya, wa/), resulting in a stimulus set of 38 speech tokens. Chinese consonant recognition was measured using a 19-alternative identification paradigm.

Each test block included 48 tokens (6 vowels \times 4 tones \times 2 speakers) for vowel/tone recognition and 38 tokens (19 consonants \times 1 tone \times 2 speakers) for consonant recognition. On each trial, a stimulus token was chosen randomly, without replacement, and presented to the subject. Following presentation of each token, the subject responded by pressing 1 of 6 buttons in the vowel test, 1 of 19 buttons in the consonant test, or 1 of 4 buttons in the tone recognition, each marked with 1 of the possible responses. The response buttons were labeled using single-vowel syllables for the vowel recognition task, consonant-/a/ context with common Chinese characters for the consonant recognition task, and 'tone 1', 'tone 2', 'tone 3', and 'tone 4' for the Chinese tone recognition task. No feedback was provided, and subjects were instructed to guess if they were not sure, although they were cautioned not to provide the same response for each guess.

Due to the generally high variability in performance among hearing-impaired children, it is difficult to separate within-subject training effects from across-subject variability. To avoid this issue, a within-subject control procedure was adopted instead of an across-subject control group procedure. The within-subject control procedure allowed 'perceptual learning' effects to be better separated from 'procedural learning' effects (i.e., task familiarization or test experience). For all subjects, baseline speech recognition performance (vowel, consonant, and tone recognition) was measured at least four times over a 2-week period. By the end of baseline testing, subjects were familiar with the test materials and procedures, reducing any 'procedural learning' effects on the subsequent training data. Asymptotic performance was observed in most subjects during the last two runs of baseline collection; thus, asymptotic performance (averaged over the last two baseline performance measures) was used as the baseline pretraining score for each subject.

Training Materials and Procedure

After baseline measures were completed, the subjects trained at home, completing targeted training tasks based on their speech recognition scores. Speech training materials included more than 1200 monosyllabic words, each spoken by 2 males and 2 females (recorded at the House Ear Institute). Custom computer-assisted speech training software, developed at the House Ear Institute and distributed by Melody Medical Instruments Corp., was used for training. The software was installed onto the subjects' home computers. For each subject, baseline recognition results were analyzed and a targeted training program was automatically generated by the computer program. For the poorest-performing subjects, training began with a 3-alternative forced-choice discrimination protocol, in which 3 sounds (labeled 'sound 1', 'sound 2', and 'sound 3') were played in sequence; 2 of the sounds were identical while the third was different. Subjects were asked to choose which sound was different. Initially, there were maximal differences (in terms of acoustic speech features) between the phonemes in the stimuli; as the subjects' performance improved (i.e., greater than 80% correct for a given level of difficulty), the differences between phonemes in the stimuli were reduced. For vowels, acoustic speech features included first and second formant frequencies (F_1 and F_2) and duration; for consonants, speech features included voice, manner and place of articulation [Miller and Nicely, 1955]. Visual feedback was provided as to whether the response was correct or wrong; auditory feedback was provided, in which the 3 sounds, now labeled with the appropriate words, were replayed in the same sequence. As the subjects progressed beyond the 3-alternative forced-choice discrimination task (performance better than 80% correct), they were trained to identify final vowels. In the identification training, only the final vowel differed between response choices, while the initial consonant and tone were kept the same; this way, subjects were better able to focus on differences between the final vowels. Initially, subjects chose between 2 responses that differed greatly in terms of speech features; as the subjects' performance improved, the difference between speech features in the response choices was reduced and/or the number of response choices was increased (up to a maximum of 6 choices) and the acoustic differences between response choices were reduced. Similar training procedures were used for consonant and tone recognition training.

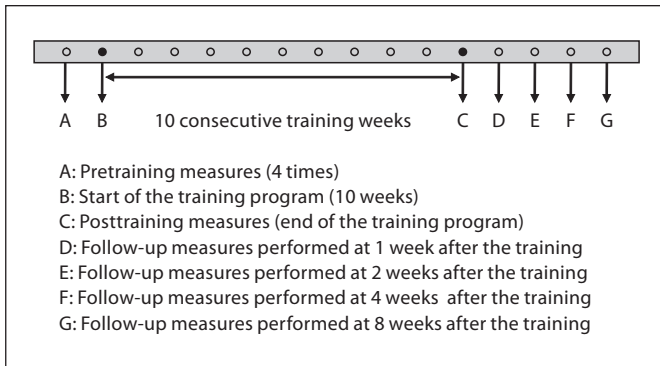


Fig. 1. Time line for testing and training.

Subjects were instructed to train for at least half an hour per day, 5 days per week, for 10 successive weeks. The training software logged each subject's training session, including the time for each exercise and the total time spent training. After completing 5 days of training each week, subjects' speech recognition performance was retested using the same speech materials as in baseline testing; recognition performances from weeks 9 and 10 were averaged as the posttraining scores for each subject. Follow-up measures of vowel, consonant and tone recognition performance were conducted 1, 2, 4 and 8 weeks after the 10-week training period ended. Figure 1 shows the time line of training and testing during the study period.

Results

Figure 2 shows individual subjects' vowel, consonant, and tone recognition scores; pretraining, posttraining, and follow-up performance is shown. Mean vowel recognition scores significantly improved from 63.1 to 84.8% correct at the end of the 10-week training period [paired t test: $t(9) = 3.575$, $p = 0.006$; power ($\alpha = 0.05$): 0.849]. The amount of improvement was highly variable, ranging from -2.0 to 58.0 percentage points; 7 out of the 10 subjects improved by 10 percentage points or more. Similarly, mean consonant recognition scores significantly improved from 39.7 to 58.2% correct [$t(9) = 7.001$, $p < 0.001$; power ($\alpha = 0.05$): 1.000]. The amount of improvement was again highly variable, ranging from 3.3 to 33.3 percentage points; 9 out of the 10 subjects improved by 10 percentage points or more. Tone recognition scores also significantly improved from 56.0 to 71.1% correct [$t(9) = 3.463$, $p = 0.007$; power ($\alpha = 0.05$): 0.801]. Again, the amount of improvement was highly variable, ranging from -3.1 to 35.4 percentage points; 6 out of the 10 subjects improved by 10 percentage points or more.

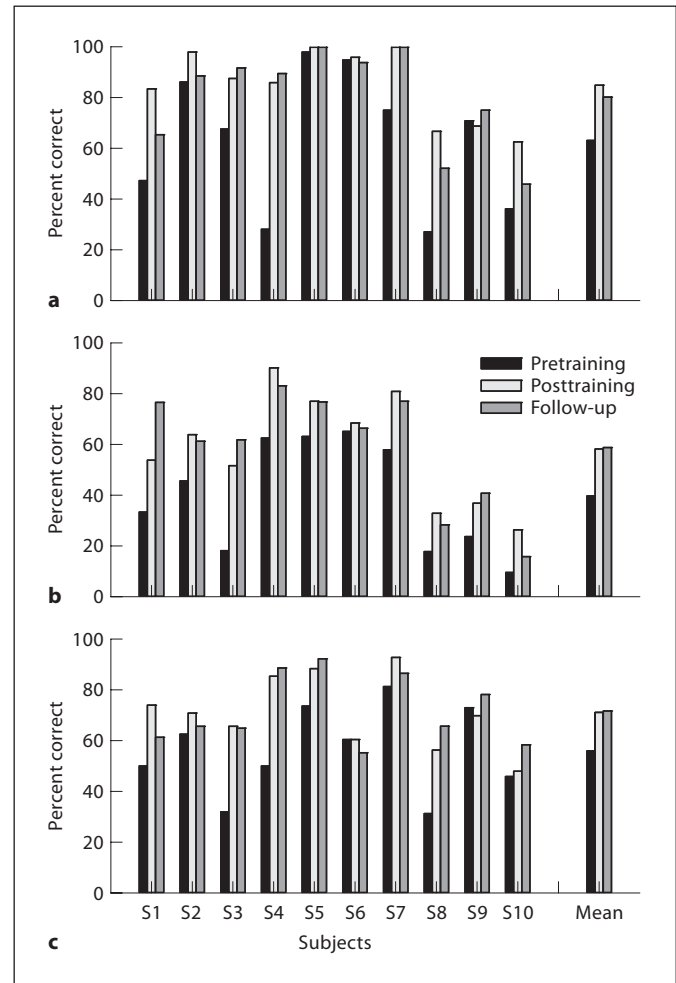


Fig. 2. Individual and mean recognition scores for pretraining, posttraining and follow-up measures. **a** Vowel recognition. **b** Consonant recognition. **c** Tone recognition.

After the 10 weeks of training were completed, vowel, consonant and tone recognition performance was remeasured at 1, 2, 4 and 8 weeks thereafter. A one-way repeated-measures ANOVA showed that there was no significant difference among these four follow-up measures for vowel [$F(3, 27) = 0.053$, $p = 0.984$], consonant [$F(3, 27) = 0.69$, $p = 0.566$] and tone recognition scores [$F(3, 27) = 0.86$, $p = 0.474$]. The four follow-up measures were averaged (fig. 2) for each subject and compared to pre- and posttraining performance. Paired t tests showed that follow-up performance was significantly higher than pretraining baseline performance for vowel [$t(9) = 2.903$, $p = 0.018$], consonant [$t(9) = 4.272$, $p = 0.002$], and tone recognition [$t(9) = 3.304$, $p = 0.009$].

Discussion

The results from the present study demonstrate that speech recognition performance of hearing-impaired children can be significantly improved with moderate amounts of daily training (e.g. half an hour per day, 5 days per week); the improved performance was largely retained 2 months after training was completed. While the results from the present study are in general agreement with those reported for adult CI users [Fu et al., 2005a], there are several interesting findings regarding the effects of auditory training in hearing-impaired children.

First, while most children benefited from the computer-assisted auditory training, there was significant variability in the amount of improvement among subjects for all speech measures. For example, the difference in vowel recognition performance after training ranged from -2.0 (S9) to 58.0 percentage points (S4), with a mean of 22 percentage points. Similarly, the improvement in consonant recognition performance ranged from 3.3 (S6) to 33.3 percentage points (S3), with a mean of 18 percentage points. Finally, the difference in tone recognition performance after training ranged from -3.1 (S9) to 35.4 percentage points (S4), with a mean of 15 percentage points. Overall, 6 out of the 10 subjects demonstrated significant improvement in all three speech measures (vowel, consonant and tone recognition). Among the 4 remaining subjects, 2 showed significant improvement in two speech measures and 1 showed significant improvement in one speech measure; 1 subject (S6) showed no improvement with training for any of the speech performance measures. Further statistical analysis revealed no significant correlation between the amount of improvement and the duration of device use, suggesting that the benefit of auditory training did not depend on experience with the hearing device. Similarly, there was no significant correlation between the amount of improvement and subject age, suggesting that auditory training outcomes did not depend on the subjects' chronological age. Interestingly, the amount of improvement was comparable between the 7 CI users and 3 HA users, even though HA users' absolute consonant recognition performance (both before and after training) was much poorer than that of CI users. The poorer consonant recognition performance was likely due to the severity of their hearing loss. The 3 HA users have either severe or profound hearing loss (mean pure-tone averages were 99, 85, and 115 dB for subjects S8, S9, and S10, respectively). Thus, they are most likely lacking the critical high-frequency information needed

for consonant recognition. Two of these subjects (S8 and S10) may be perfect candidates for cochlear implantation. It is not surprising that these subjects do not benefit too much from their hearing aids, given the severity of their hearing loss.

Second, while previous studies have shown that auditory training can significantly improve recognition of sentences and/or segmental information, the present study demonstrates that moderate amounts of auditory training can also improve recognition of suprasegmental information (e.g. Chinese tones). Previous studies have shown that temporal envelope cues (e.g. amplitude contour and periodicity fluctuations) contribute most strongly to the moderate levels of tone recognition by CI patients [Fu et al., 1998, 2004; Fu and Zeng, 2000; Hsu et al., 2000; Luo and Fu, 2004]. The results from the present study suggest that reception of temporal envelope cues can be improved with moderate auditory training. However, the degree of improvement with training may be highly patient-dependent, as illustrated by the variability in improvement among the participating subjects.

Third, the results from the present study suggest that a computer-based auditory training tool, such as the one used in the present study, may be a useful alternative or complement to auditory rehabilitation provided by clinicians. For hearing-impaired children, auditory rehabilitation is a critical process for hearing and speech development. Unfortunately, access to auditory rehabilitation provided by hospitals and hearing clinics may be limited for many hearing-impaired children, due to time constraints, expense and the proximity of patients to the rehabilitation site. When provided with user-friendly speech training software, hearing-impaired children can easily perform self-guided auditory rehabilitation at home; as a result, the time and cost of site-specific auditory rehabilitation can be reduced.

While results from this and previous training studies [Fu et al., 2005a] suggest great promise for auditory training in CI patients, many challenges remain in terms of maximizing the benefit of training. Many factors may affect training outcomes [Fu et al., 2005b]. One such factor is the training protocol. Fu et al. [2005b] have recently evaluated the effects of four different protocols on training outcomes in normal-hearing subjects listening to spectrally shifted vowels. The 'test-only' protocol provided an experimental control for procedural learning. The 'preview' protocol (direct preview of the test stimuli with novel talkers, immediately before testing) and the 'vowel contrast' protocol (extended training with novel mono-

syllabic words, spoken by novel talkers, targeting acoustic differences between medial vowels) were more bottom-up in approach, as subjects were trained to identify only small acoustic differences between stimuli (without linguistic or contextual cues). The 'sentence training' protocol (modified connected discourse) was more top-down in approach, as subjects were provided with contextual cues in the sentence stimuli. With the test-only and sentence training protocols, there was no significant change in vowel recognition during the 5-day training period. However, vowel recognition significantly improved after the 5-day training with either the preview or vowel contrast training protocol.

Besides different protocols, different training materials (i.e., speech stimuli) may also influence training outcomes. A variety of speech materials have been used in previous auditory training studies (e.g. natural speech, synthetic speech, noisy speech, acoustically modified speech). Differences in CI patients' initial performance may dictate which training materials may be most appropriate. For poor performers (for whom speech recognition in noise is nearly impossible), phonemic contrast training may best improve speech performance in quiet. For moderate-to-good performers, speech training in noise (whether with sentences or phonemes) may improve performance in noise and in quiet, depending on initial performance levels. Of course, it is most desirable

to identify the training materials that will be most effective and most easily generalized to a variety of listening conditions.

Conclusion

Nine out of 10 pediatric subjects were able to significantly improve performance in one or more speech measures after 10 weeks of computer-assisted auditory training, conducted at subjects' homes with minimal supervision. There was significant intersubject variability in terms of the amount of improvement after training. Improved performance was largely retained 2 months after training was completed. The results suggest that moderate amounts of auditory training using a computer-based rehabilitation tool can be a convenient and cost-effective approach to improve the speech perception performance of hearing-impaired children.

Acknowledgements

We are grateful to all research participants for their considerable time spent with this experiment. We would also like to thank John J. Galvin 3rd for editing assistance and four anonymous reviewers for the useful comments and suggestions. The research was partially supported by NIDCD grant R01-DC004792.

References

- Busby PA, Roberts SA, Tong YC, Clark GM: Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant. *Br J Audiol* 1991;25:291-302.
- Dawson PW, Clark GM: Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant. *Ear Hear* 1997;18:488-501.
- Fu Q-J, Galvin JJ 3rd, Wang X, Nogaki G: Moderate auditory training can improve speech performance of adult cochlear implant users. *J Acoust Soc Am (ARLO)* 2005a;6:106-111.
- Fu Q-J, Hsu C-J, Horng M-J: Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant patients. *Ear Hear* 2004;25:501-508.
- Fu Q-J, Nogaki G, Galvin JJ 3rd: Auditory training with spectrally shifted speech: an implication for cochlear implant users' auditory rehabilitation. *J Assoc Res Otolaryngol* 2005b;6:180-189.
- Fu Q-J, Zeng F-G, Shannon RV, Soli SD: Importance of tonal envelope cues in Chinese speech recognition. *J Acoust Soc Am* 1998;104:505-510.
- Fu Q-J, Zeng F-G: Effects of envelope cues on Mandarin Chinese tone recognition. *Asia Pac J Speech Lang Hear* 2000;5:45-57.
- Hesse G, Nelting M, Mohrmann B, Laubert A, Ptok M: Intensive inpatient therapy of auditory processing and perceptual disorders in childhood. *HNO* 2001;49:636-641.
- Hsu CJ, Horng MJ, Fu Q-J: Effects of the number of active electrodes on tone and speech perception by Nucleus-22 cochlear implant users with the SPEAK strategy. *Adv Otorhinolaryngol* 2000;57:257-259.
- Lin M-C: The acoustic characteristics and perceptual cues of tones in Standard Chinese. *Chinese Yuwen* 1988;204:182-193.
- Luo X, Fu Q-J: Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. *J Acoust Soc Am* 2004;116:3659-3667.
- Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL, Tallal P: Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* 1996;271:77-81.
- Miller GA, Nicely PE: An analysis of perceptual confusion among some English consonants. *J Acoust Soc Am* 1955;27:338-352.
- Sweetow R, Palmer CV: Efficacy of individual auditory training in adults: a systematic review of the evidence. *J Am Acad Audiol* 2005;16:494-504.
- Zeng F-G: Trends in cochlear implants. *Trends Amplif* 2004;8:1-34.