Bimodal audio–visual training enhances auditory adaptation process

Tetsuaki Kawase[a,b,c], Shuichi Sakamoto[d], Yoko Hori[c], Atsuko Maki[c], Yôiti Suzuki[d] and Toshimitsu Kobayashi[c]

Effects of auditory training with bimodal audio–visual stimuli on monomodal aural speech intelligibility were examined in individuals with normal hearing using highly degraded noise-vocoded speech sound. Visual cue simultaneously presented with auditory stimuli during the training session significantly improved auditory speech intelligibility not only for words used in the training session, but also untrained words, when compared with the auditory training using only auditory stimuli. Visual information is generally considered to complement insufficient speech information conveyed by the auditory system during audio–visual speech perception. However, the present results showed another beneficial effect of audio–visual training that the visual cue enhances the auditory adaptation process to the degraded new speech sound, which is different from those given during bimodal training. *NeuroReport* 20:1231–1234 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins.

NeuroReport 2009, **20**:1231–1234

Keywords: audio–visual, auditory rehabilitation, multimodality

[a]Laboratory of Rehabilitative Auditory Science, Tohoku University Graduate School of Biomedical Engineering, Departments of [b]Audiology, [c]Otolaryngology – Head and Neck Surgery, Tohoku University Graduate School of Medicine and [d]Research Institute of Electrical Communication, Tohoku University, Sendai, Japan

Correspondence to Dr Tetsuaki Kawase, MD, Laboratory of Rehabilitative Auditory Science, Tohoku University Graduate School of Biomedical Engineering, Sendai 980-8574, Japan
Tel: +81 22 717 7303; fax: +81 22 717 7307; e-mail: kaw@m.tains.tohoku.ac.jp

Received 2 May 2009 **accepted** 16 June 2009

Introduction

The human brain effectively integrates information from multisensory modalities during the perception of external signals. This multimodal processing is beneficial for fast and accurate cognition of information. Deteriorated speech communication in the presence of degraded auditory conditions, such as background noise and in patients with hearing loss, is improved by the combined presentation of visual speech [1–3]. If the degraded speech is perceived as bimodal audio–visual stimuli, visual information from the speaker's face can be effectively utilized to make up for inadequate auditory information [4–6].

In contrast, it is unclear whether bimodal speech perception is beneficial for the monomodal aural rehabilitative process to compensate for the degraded auditory conditions occurring in patients with cochlear implants, which are sensory prostheses intended to restore hearing to deafened patients by electric stimulation of the remnant auditory nerve. The input acoustic signal restored by the cochlear implant is spectrally compressed relative to the normal tonotopic pattern and contains limited temporal information for the perception of pitch [7]. Therefore, even postlingually deafened patients usually need a certain period of time to adapt to the modified properties of the input acoustic signal ('rehabilitative process').

Better monomodal speech intelligibility after cochlear implantation in patients with visual disturbance [8,9] supports the idea that monomodal speech training (i.e. auditory-only) is advantageous for aural rehabilitation. In contrast, several recent reports have shown that bimodal auditory–visual training facilitates the monomodal visual learning process [10,11], suggesting that bimodal audio–visual stimuli not only facilitate the perceptual process for deteriorated speech, but also improve the monomodal aural rehabilitative process to compensate for severely degraded auditory conditions [12].

To clarify whether bimodal audio–visual training can really facilitate the aural rehabilitative process more than monomodal speech training or not, in this study, the effects of auditory training with bimodal audio–visual stimuli on the monomodal aural adaptation process were examined in individuals with normal hearing using highly degraded noise-vocoded speech sound (NVSS), which is often used as a simulation of cochlear implant speech [13,14]. This speech sound is hardly intelligible at first listening. However, adequate auditory training can improve the intelligibility of NVSS.

Materials and methods

This study included 34 normal volunteers (22 males and 12 females, mean age 26.4 years) with normal hearing without histories of auditory diseases or neurological

0959-4965 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0b013e32832fbef8

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

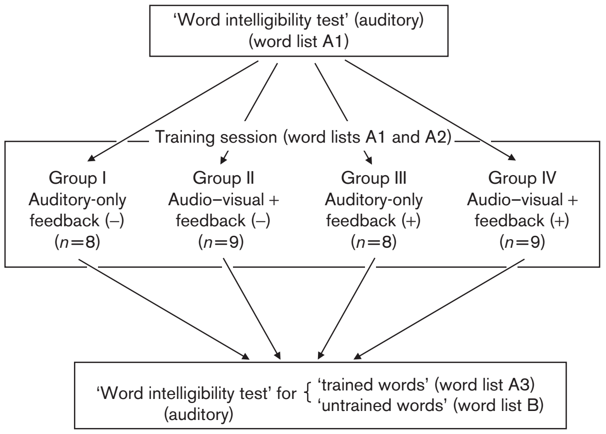
disorders. Audio–visual recordings were made of four lists of 50 four-mora Japanese words spoken by a Japanese female trained speaker. Three word lists (lists A1, A2, and A3) consisted of the same words in different orders. The other word list (list B) contained completely different words. The word lists were selected from phonetically balanced and familiarity-controlled Japanese word lists for the spoken-word intelligibility test (Familiarity-controlled Word Lists 2003: FW03) [15,16]. FW03 consists of four word-familiarity ranks (lower, lower middle, upper middle, and high familiarities), and the recorded word lists can be obtained from Speech Resources Consortium (*http://research.nii.ac.jp/src/eng/index.html*). Words of upper-middle familiarity were used in this study.

To simulate the speech signal provided by cochlear implants as heard by patients with severe deafness, highly degraded NVSS was generated as follows: first, the original speech sound was band-pass filtered into two frequency bands (from 2667 to 3333 Hz and from 3333.4 to 4000.0 Hz) by two types of band-pass filter. The amplitude envelope of the output of each frequency band was extracted by the Hilbert transform. Then, the amplitude envelopes of the two frequency bands were multiplied by narrow-band noise with two frequency bands (from 3111 to 3333 Hz and from 3555 to 4000 Hz), respectively, to compress the signal spectrally. The maximum amplitude of the band noises was normalized. These conditions are regarded to simulate the supply of the high-frequency component of the speech information to two active cochlear electrodes with a certain space, which could be implanted only at the very basal part of the cochlear nerve.

After the initial assessment of auditory speech intelligibility (no visual cue) using word list A1, the participants were divided into four groups. As shown in Fig. 1, these four groups are divided by undergoing different training sessions with word lists A1 and A2 and combinations of presence/absence of visual cue and presence/absence of feedback (Fig. 1). In training sessions, two word lists (A1 and A2) consisting of the same 50 four-mora words in different orders were alternately presented 10 times (five times each). The participants were instructed to write down what was heard after the presentation of each word to focus attention on the training stimuli as far as possible and to minimize differences in the attention level to the stimuli between the different training conditions. The effects of these different training sessions on auditory speech intelligibility (no visual cue) were assessed for trained words (list A3) and for untrained words (list B) after the training session.

All auditory and audio–visual speech stimulations were presented using a 25-inch TV monitor and its built-in speaker, which was set 1.5 m in front of the participant. When only auditory speech was presented to the participants, the TV screen was covered with a large cardboard screen. Sound pressure level was adjusted to 80 dB L_{Aeq} in all tests as well as training conditions. All procedures (speech intelligibility test before and after auditory training as well as the training session) were conducted serially within 3 h except for a short rest.

Fig. 1



Schema of the experimental procedure. After the initial measurement of speech intelligibility for word list A1, participants were divided into four groups with different training conditions: auditory-only without feedback (group I, six males and two females), audio–visual without feedback (group II, seven males and two females), auditory-only with feedback (group III, five males and three females), and audio–visual with feedback (group IV, four males and five females). In the training session, 50 words of lists A1 and A2 were alternately presented 10 times (five times each). The participants were instructed to write down what was heard after the presentation of each word. The correct answer was provided as feedback after the response in groups III and IV. After the training session using word lists A1 and A2, the speech intelligibility test was conducted using trained word list A3 and untrained word list B.

All parts of this study were approved by the Ethics Committee of Tohoku University School of Medicine, and were performed in accordance with the guidelines of the Declaration of Helsinki.

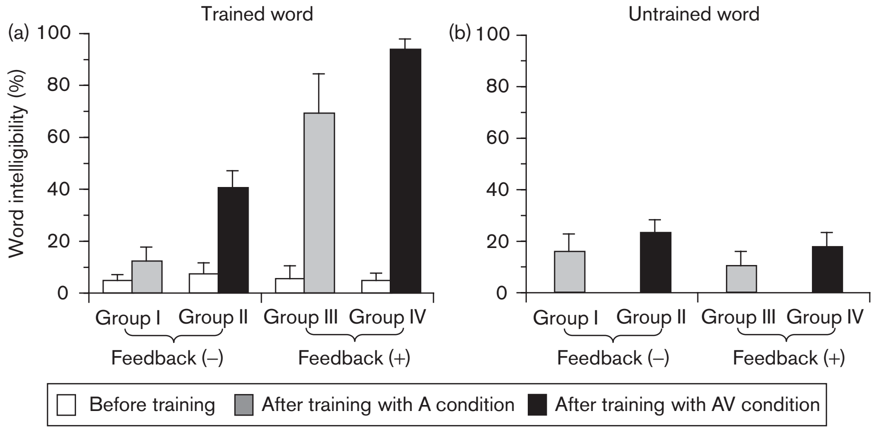
Results

The speech intelligibility was significantly improved in all the four groups but was significantly different between the different training conditions (Fig. 2).

Results were analyzed by using three-way analysis of variance; the three factors were visual cue (presence or absence during training session), word list (trained word list or untrained word list), and feedback (with feedback or without feedback). There were significant interactions between visual cue and word list [F(1,30) = 37.68, $P < 0.01$] and between word list and feedback [F(1,30) = 383.41, $P < 0.01$]. The simple main effect of visual cue was significant for both word lists ($P < 0.01$);

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Fig. 2



Word intelligibilities for (a) the trained word list (before and after training) and (b) the untrained word list (after training) in different four training conditions (see text for further details).

that is, the visual cue simultaneously presented with auditory stimuli during the training session significantly improved the auditory speech intelligibility not only for trained words but also untrained words, irrespective of the feedback condition. The simple main effect of feedback was also significant for both word lists (trained word list: $P < 0.01$; untrained word list: $P < 0.05$). The presence of feedback during the training session resulted in significantly better speech intelligibility for trained words (Fig. 2a). In contrast, use of feedback resulted in lower scores than those without feedback in the posttraining test for untrained words (Fig. 2b), showing overtraining effects. Facilitative visual effects on posttraining auditory performance were also observed regardless of the overtraining effects.

These results indicate that combined audio–visual training is beneficial on the monomodal auditory adaptation process not only for trained words but also untrained words.

Discussion

Visual information is generally considered to complement insufficient speech information conveyed by the auditory system during audio–visual speech perception. However, the present results showed another beneficial effect of audio–visual training that the visual cue enhances the auditory adaptation process to the degraded new speech sound during bimodal training. The present results are important not only in the clinical aspects, such as auditory rehabilitation of patients with cochlear implant, but also as another aspect of tightly coupled audio–visual multimodal sensory interaction in the brain; that is, the audio–visual interaction can be observed both as the perceptual process of multimodality and also as the after-effect of multimodal training even with short-term training and even for untrained words.

The after-effect of recurrent exposure to audio–visual stimuli has been also investigated using the phonemic discriminative task for incongruent audiovisual speech and for second-language learners [17,18]. Two-forced judgment of the perception of several ambiguous sounds intermediate between /aba/ and /ada/ (place-of-articulation auditory continuum), that is, what is heard must be chosen from /aba/ or /ada/, could be affected (recalibrated) after recurrent exposure to the ambiguous sounds with visual speech information articulating either /aba/ or /ada/ [17]. In contrast, a recent study of the effect of audio–visual bimodal training on the phonemic discriminative task with ordinary speech sounds in second-language learners hints that the bimodal training effect was different depending on the visual cues to phonemic contrast [18]. Training with audio–visual cues was more effective than training with only audio cues, if the visual cues to phonemic contrast were sufficiently salient, such as the /v/-/b/-/p/ labial/labiodental contrast, whereas bimodal audio–visual training was not effective in the discriminative task for the /l/–/r/ contrast [18]. Therefore, based on these two studies, the after-effect of bimodal training might be observed more clearly under the condition that a larger lip reading effect is expected, such as sufficiently salient visual cues to phonemic contrast. This finding seems to support the idea that the visual information obtained during the audio–visual bimodal training is strongly related to the 'after-effect' phenomenon, although the possible involvement of other factors, such as the different attention levels caused by the addition of visual stimuli, could not be excluded.

Moreover, patients with deafness have better audio–visual integration, such as lip reading, than normal listeners [12], and so might use such information more effectively. In this sense, greater effects of bimodal

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

audio–visual training might be obtained in the auditory rehabilitation process of patients with auditory prosthesis than those obtained in this study of participants with normal hearing.

Audio–visual integration in impaired listeners has been an important topic relating to multimodal sensory interaction in the brain, but most studies have focused on the classical audio–visual interaction as lip reading [4,5,12,19,20]. The effectiveness of audio–visual bimodal training in the actual rehabilitative process in patients with auditory prosthesis has not yet been confirmed, but is likely based on the positive results of bimodal training effects on monomodal adaptation or the learning process seen in normal individuals (including the present findings) [10,11,17,18]. However, how audio–visual training can affect the actual rehabilitative process should be investigated, considering the difference in time scale between the actual rehabilitative process and the adaptation or learning processes observed so far.

Conclusion

The visual cue simultaneously presented with auditory stimuli enhances the auditory adaptation process to the degraded new speech sound during audio–visual bimodal training. The present results are important not only in the clinical aspects, such as auditory rehabilitation of patients with cochlear implants, but also as another insight into the tightly coupled audio–visual multimodal sensory interaction in the brain.

Acknowledgements

The authors thank Shun Numahata for his assistance in processing the speech materials. This study was supported by a grant from the Ministry of Education, Science and Culture [Grant-in-Aid for Scientific Research (C) 19591954].

References

1 Rosen SM, Fourcin AJ, Moore BC. Voice pitch as an aid to lipreading. *Nature* 1981; **291**:150–152.

2 Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 1954; **26**:212–215.

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

4 Giraud AL, Truy E. The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia* 2002; **40**:1562–1569.

5 Kawase T, Yamaguchi K, Ogawa T, Suzuki K, Suzuki M, Itoh M, *et al.* T. Recruitment of fusiform face area associated with listening to degraded speech sounds in auditory-visual speech perception: a PET study. *Neurosci Lett* 2005; **382**:254–258.

6 Sekiyama K, Kanno I, Miura S, Sugita Y. Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res* 2003; **47**:277–287.

7 Shannon RV. Understanding hearing through deafness. *Proc Natl Acad Sci U S A* 2007; **104**:6883–6884.

8 El-Kashlan HK, Boerst A, Telian SA. Multichannel cochlear implantation in visually impaired patients. *Otol Neurotol* 2001; **22**:53–56.

9 Saeed SR, Ramsden RT, Axon PR. Cochlear implantation in the deaf-blind. *Am J Otol* 1998; **19**:774–777.

10 Frassinetti F, Bolognini N, Bottari D, Bonora A, Làdavas EJ. Audiovisual integration in patients with visual deficit. *J Cogn Neurosci* 2005; **17**:1442–1452.

11 Seitz AR, Kim RS, Shams L. Sound facilitates visual learning. *Curr Biol* 2006; **16**:1422–1427.

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

13 Fu QJ, Nogaki G, Galvin JJ III. Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. *Assoc Res Otolaryngol* 2005; **6**:180–189.

14 Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995; **270**:303–304.

15 Amano S, Sakamoto S, Kondo T, Suzuki Y. Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese. *Speech Commun* 2009; **51**:76–82.

16 Sakamoto S, Suzuki Y, Amano S, Ozawa K, Kondo T, Sone T. New lists for word intelligibility test based on word familiarity and phonetic balance. *J Acoust Soc Jpn* 1998; **54**:842–849.

17 Bertelson P, Vroomen J, de Gelder B. Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychol Sci* 2003; **14**:592–597.

18 Hazan V, Sennema A, Iba M, Faulkner A. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Commun* 2005; **47**:360–378.

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

20 Schorr EA, Fox NA, van Wassenhove V, Knudsen EI. Auditory-visual fusion in speech perception in children with cochlear implants. *Proc Natl Acad Sci U S A* 2005; **102**:18748–18750.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.