# Chapter 10
# Importance of Visual Cues in Hearing Restoration by Auditory Prosthesis

**Tetsuaki Kawase, Yoko Hori, Takenori Ogawa, Shuichi Sakamoto, Yôiti Suzuki, and Yukio Katori**

**Abstract** Auditory prostheses, such as cochlear implant and auditory brainstem implant, are used clinically to restore the hearing of patients with sensorineural hearing loss. These devices can considerably improve the auditory information conveyed to the auditory cortex, but proper rehabilitation process is usually necessary to restore auditory communication to an adequate level. Therefore, improvements in the auditory information provided by the prosthesis can be complemented by better rehabilitation process.

Moreover, the complementary role of visual cues is also important. The "lip-reading" phenomenon is well known in patients with degraded speech perception; i.e., reduced speech perception in the presence of poor auditory conditions, such as background noise and in patients with hearing loss, is improved by the combined presentation of visual speech. In addition to such conventional lip-reading, audio-visual speech has another beneficial role in the auditory rehabilitation process; i.e., the visual cue enhances the auditory adaptation process to the degraded speech sound.

In the present paper, these two aspects of audio-visual speech in auditory rehabilitation are reviewed.

**Keywords** Audio-visual speech • Auditory prosthesis • Lip-reading • Rehabilitation

T. Kawase (✉)
Laboratory of Rehabilitative Auditory Science, Tohoku University Graduate School of Biomedical Engineering, Sendai 980-8574, Japan

Department of Audiology, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan

Department of Otolaryngology-Head and Neck Surgery, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan
e-mail: kawase@orl.med.tohoku.ac.jp

Y. Hori • T. Ogawa • Y. Katori
Department of Otolaryngology-Head and Neck Surgery, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan

S. Sakamoto • Y. Suzuki
Research Institute of Electrical Communication, Tohoku University, Sendai 980-8574, Japan



K. Sasaki et al. (eds.), *Interface Oral Health Science 2014*,
DOI 10.1007/978-4-431-55192-8_10

## 10.1 Introduction

Sound vibrations in the air are transmitted to the inner ear via the ear drum and ossicular chain. The hair cell system, located in the inner ear, converts the sound vibrations into the electrical spike signals of the cochlear nerves. Therefore, the hair cell system is very important in the mechanism of electrical transduction in the inner ear. However, this important transducer system never regenerates after damage. Consequently, various types of auditory prostheses have been developed to restore the hearing of patients with sensorineural hearing loss.

Generally speaking, auditory prostheses are classifiable into two types depending on the fundamental concept of the device: some devices such as hearing aids (HAs), bone-anchored hearing aids (BAHAs), and Vibrant SoundBridge middle ear implants (MEIs) increase the energy of the sound vibrations transmitted to the damaged inner ear; whereas other devices such as cochlear implants (CIs), auditory brainstem implants (ABIs), and auditory midbrain implants (AMIs) stimulate the auditory system electrically (Fig. 10.1). Usually these latter devices (CI, ABI, AMI, etc.) are used if the hearing loss is too severe to use the former

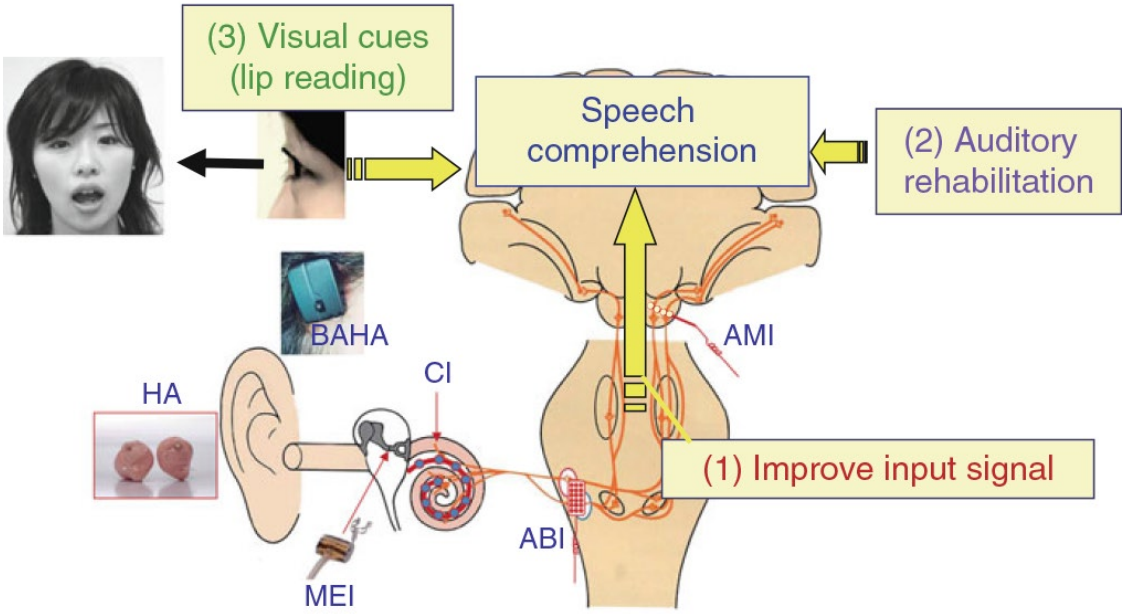


**Fig. 10.1** Basic strategy for better speech intelligibility in auditory rehabilitation. Devices such as the hearing aid (HA), bone-anchored hearing aid (BAHA, a type of HA based on bone conduction), and middle ear implant (MEI, a direct-drive, implantable middle ear device, which mechanically stimulates the ossicles, mimicking the natural hearing process) increase the energy of the sound vibrations transmitted to the damaged inner ear. In contrast, the cochlear implant (CI), auditory brainstem implant (ABI), and auditory midbrain implant (AMI) stimulate the auditory system electrically. The CI restores hearing by direct electrical stimulation of the cochlear nerve, whereas the ABI and AMI directly stimulate the auditory pathway at the cochlear nucleus and mid-inferior colliculus, respectively. Auditory information can be considerably improved by these electrical stimulation devices, but is usually insufficient. Therefore, rehabilitation and sometimes visual information known as the lip-reading effect are usually necessary to restore auditory communication to an adequate level

devices (HA, BAHA, MEI, etc.). All these devices can considerably improve auditory information conveyed to the auditory cortex, even in patients with severely degraded hearing loss requiring CI and/or ABI, however, a suitable rehabilitation process is usually necessary to restore a certain level of auditory communication. The basic strategy for improving speech intelligibility by auditory rehabilitation is presented in Fig. 10.1.

Therefore, it is important to improve both the quality of the auditory information that can be provided by each prosthesis ((1) in Fig. 10.1), and the rehabilitation process for individual patients ((2) in Fig. 10.1). Moreover, the complementary role of visual cues is also important ((3) in Fig. 10.1). The "lip-reading" phenomenon is well known in patients with degraded speech perception; i.e., reduced speech perception in the presence of poor auditory conditions, such as background noise and in patients with hearing loss, is improved by the combined presentation of visual speech [12, 15]. If the degraded speech can be perceived as bimodal audio-visual stimuli, the visual information from the speaker's face can be effectively utilized to compensate for the inadequate auditory information [2, 9, 13]. In addition to such conventional lip-reading, audio-visual speech has another beneficial role in the auditory rehabilitation process; i.e., the visual cue enhances the auditory adaptation process to the degraded speech sound [10].

Here, these two aspects of audio-visual speech in auditory rehabilitation are reviewed.

## 10.2 Recruitment of Visual Cues in Degraded Speech Conditions

Perception of external signals is followed by integration of the information from multisensory modalities in the brain. Such multi-modal processing results in fast and accurate recognition of the perceived signals. Speech perception effectively utilizes the visual information from the speaker's face not only in patients with hearing loss but also in healthy subjects; i.e., speech perception in degraded conditions such as background noise can be improved by visual information obtained from the speaker's face [12, 15]. Therefore, visual cues (speaker's face) presented with auditory cues (speech sound) will be utilized to complement the auditory information in every situation. However, the degree of recruitment of visual cues will depend on the degree of deterioration of speech perception [9].

Positron emission tomography (PET) was used to evaluate the effect of this recruitment of visual cues on the activation of additional brain areas caused by degradation of auditory input, as presented in Fig. 10.2. This PET study compared brain activation caused by the presentation of a visual cue (facial movement at speech) with control conditions (visual noise) under two different audio-conditions, normal speech and degraded speech. Lip-reading for degraded speech caused more activations than for normal speech in V2 and V3 of visual cortex as well as in the right fusiform gyrus of the temporal lobe (see [9] for details). The right fusiform

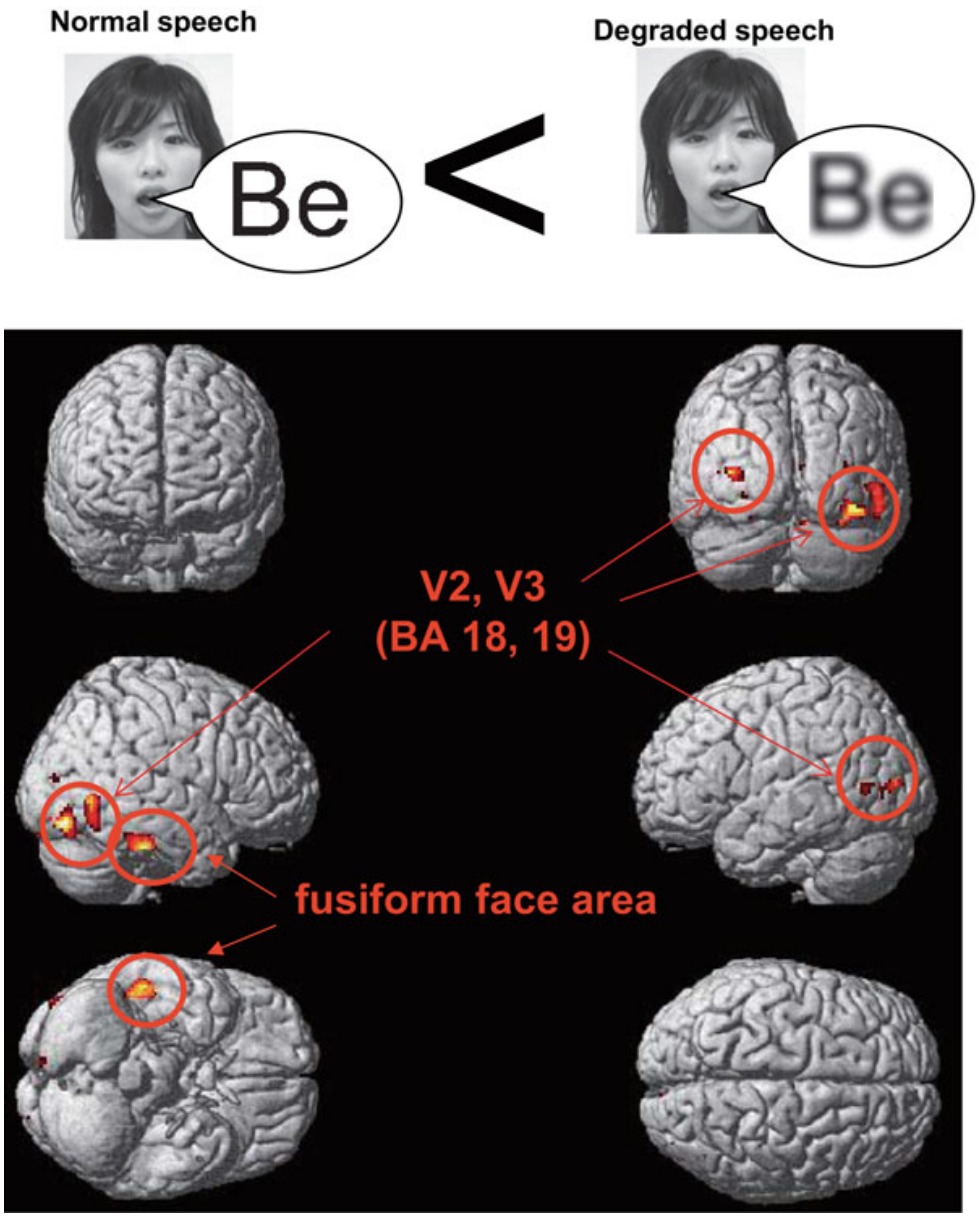


**Fig. 10.2** Additional recruitment of brain areas caused by degradation of auditory input (unpublished figure using our data published previously [9]). Positron emission tomography (PET) was used to compare brain activation caused by the presentation of a visual cue (facial movement at speech) with control conditions (visual noise) under two different audio-conditions, normal speech and degraded speech. Significant brain activation is presented during lip-reading under degraded speech compared with normal speech. Suprathreshold voxels ($P < 0.001$, uncorrected for multiple comparisons, $k > 20$ voxels) superimposed on the 3D-rendered surface image. Lip-reading for degraded speech caused more activations than for normal speech in V2 and V3 of visual cortex as well as in the right fusiform gyrus of the temporal lobe (see Kawase et al. 2005 [9] for details)

gyrus of the temporal lobe is a well-known brain area known as the fusiform face area (FFA). The FFA, together with the inferior occipital gyri and the superior temporal sulcus, is one of the three important brain regions in the occipitotemporal visual extrastriate cortex related to human face perception [3–8, 11]. Therefore,

activation of the FFA during auditory-visual speech perception is very likely. The present study indicated that the degree of activation of FFA depends on the degree of the degradation of auditory cues. This observation is consistent with the hypothesis that more visual information than usual is recruited under conditions of degraded auditory information.

## 10.3 Auditory Training with Bimodal Audio-Visual Stimuli

These investigations of perception of bimodal audio-visual stimuli under degraded speech conditions show that visual information from the speaker's face can be effectively utilized to make up for inadequate auditory information. Therefore, combined presentation of visual speech information is important in speech communication in the presence of degraded auditory conditions, such as background noise and in patients with hearing loss.

On the other hand, audio-visual speech cues have another beneficial role in the auditory rehabilitation process; i.e., the visual cue enhances the auditory adaptation process to the degraded speech sound [10]. In that study, auditory training was examined in normal volunteers using highly degraded noise-vocoded speech sound (NVSS), which is often used as a simulation of the effects of cochlear implant on speech [1, 14]. NVSS is hardly intelligible at first listening, but adequate auditory training can improve the intelligibility of NVSS. After the initial assessment of auditory speech intelligibility (no visual cue), the subject underwent different training sessions with combinations of presence/absence of visual cue and presence/absence of feedback of the correct answer. The training sessions used two word lists consisting of the same 50 four-mora words in different orders which were alternately presented ten times (five times each). The effects of these different training sessions on auditory speech intelligibility (no visual cue) were assessed for the trained words (in different order from those used in the training session) as well as untrained words after the training session (see [10] for details).

The effects of the presence of visual cues during the training session on word intelligibility after the training session are presented in Fig. 10.3 (feedback (−) groups) and 4 (feedback (+) groups). Speech intelligibility after the training session was significantly improved in all training groups but was significantly different between the different training conditions. Visual cues simultaneously presented with auditory stimuli during the training session significantly improved auditory speech intelligibility compared to only auditory stimuli. Feedback during the training session also resulted in significantly better speech intelligibility for trained words (Fig. 10.4). In contrast, feedback resulted in lower scores compared to without feedback in the post-training test for untrained words (Fig. 10.4), showing over-training effects. However, facilitative visual effects on post-training auditory performance were also observed regardless of the over-training effects. These results indicate that combined audio-visual training has beneficial effects in auditory rehabilitation.

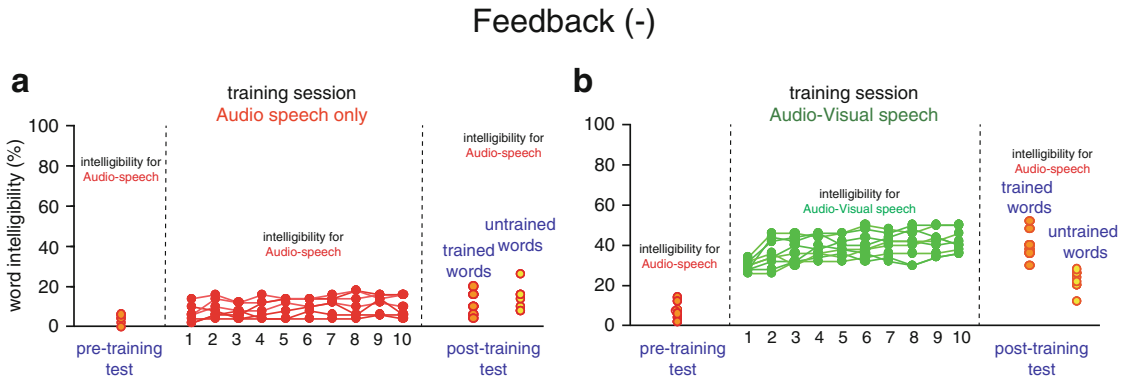


**Fig. 10.3** Effects of the presence of visual cues during the training session on word intelligibility after the training session (no feedback condition) (unpublished figure using our data published previously [10]). (**a**) Training with auditory cue only (without visual cues), (**b**) training with auditory + visual cues (lip-reading condition). Word intelligibilities (no visual cue) before and after training are presented along with those during the training (learning curves). Intelligibilities during the training are those for training modalities; i.e. for only auditory speech (**a**) and audio-visual speech (**b**), respectively. Intelligibilities after training are shown for trained words (intelligibility for words used in the training session) and for untrained words (intelligibility for words not used in the training session)

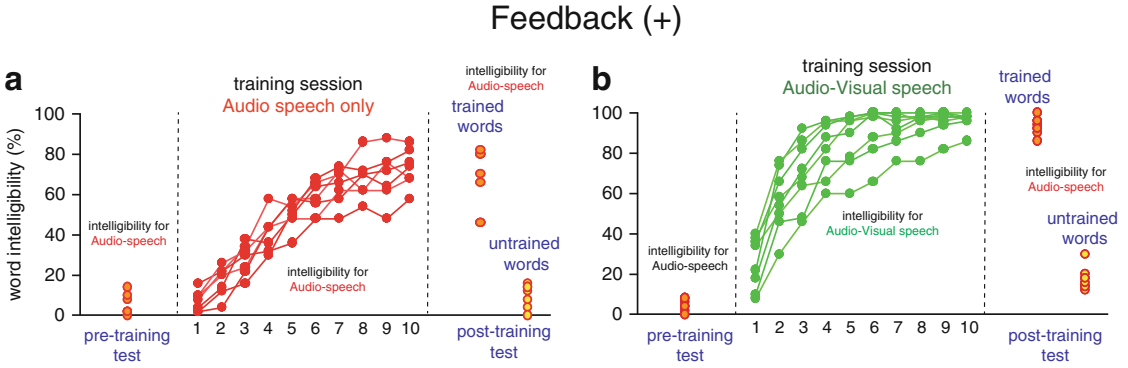


**Fig. 10.4** Effects of the presence of visual cues during the training session on word intelligibility after the training session (with feedback condition) (unpublished figure using our data published previously [10]). (**a**) Training with auditory cue only (without visual cues), (**b**) training with auditory + visual cues (lip-reading condition). Word intelligibilities (no visual cue) before and after training are presented along with those during the training (learning curves). Intelligibilities during the training are those for training modalities; i.e. for only auditory speech (**a**) and audio-visual speech (**b**), respectively. Intelligibilities after training are shown for trained words (intelligibility for words used in the training session) and for untrained words (intelligibility for words not used in the training session). In contrast with feedback (−) groups in Fig. 10.3, better speech intelligibilities were obtained for trained words. On the other hand, feedback resulted in lower scores compared to without feedback in the post-training test for untrained words, showing over-training effects. Facilitative visual effects on post-training auditory performance were also observed regardless of the over-training effects

The effects of different training sessions on the intelligibility are presented in Fig. 10.5, divided into each "mora" of the words. Basically, similar trends to those found based on word intelligibility were also observed by this "mora"-based analysis, although the intelligibility was different for the first, second, third, and fourth moras.

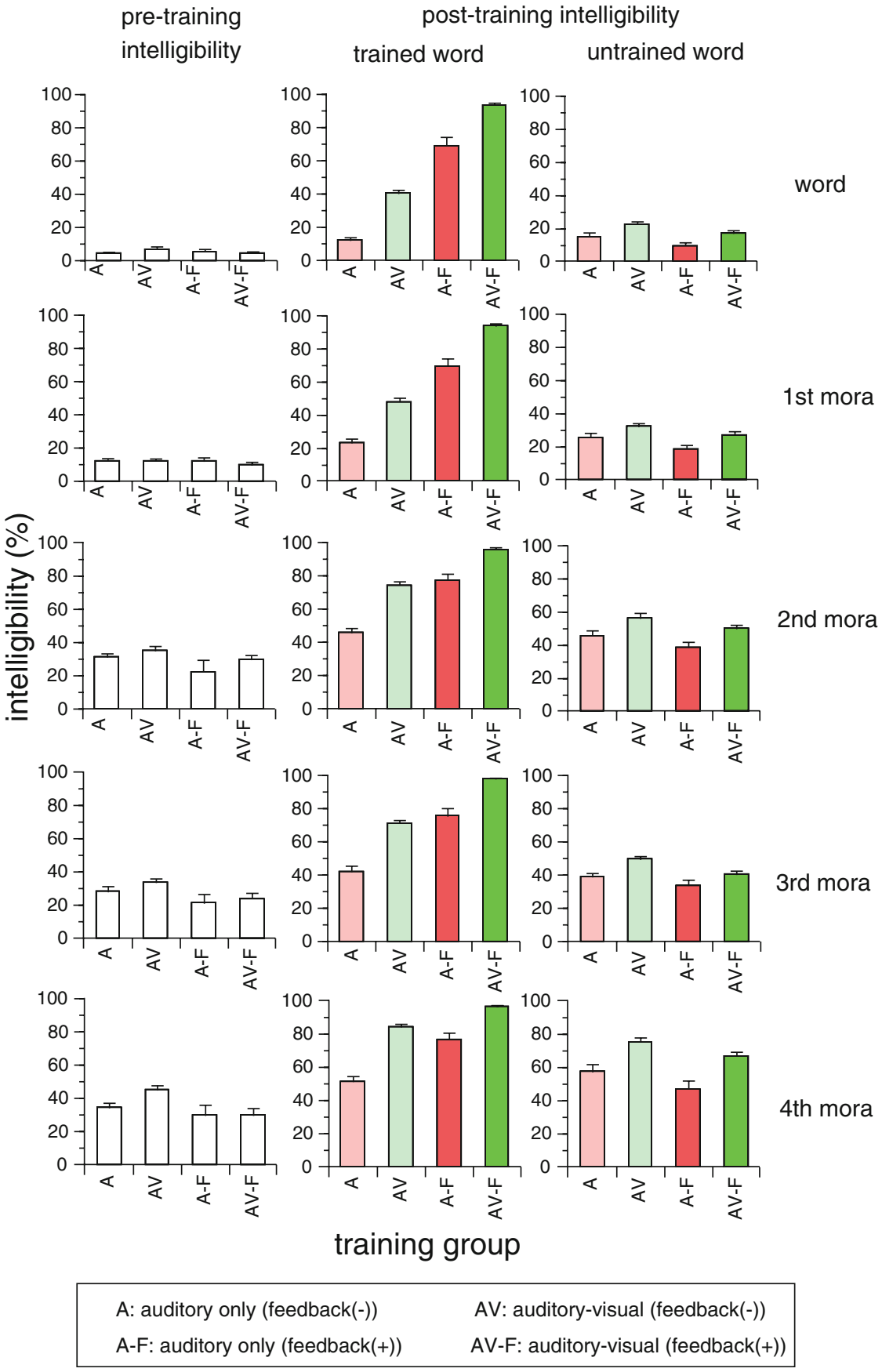


**Fig. 10.5** Effects of different training sessions on "mora" intelligibilities (unpublished figure using our data published previously [10]). In addition to word intelligibilities, the intelligibilities for the first, second, third, and fourth moras are presented for four training sessions. Although the intelligibilities for the first mora were significantly smaller, basically similar trends to those found based on word intelligibility were also observed

Visual information is generally considered to complement insufficient speech information in speech comprehension. However, the present results revealed another beneficial effect of audio-visual training; i.e., the visual cue enhances the auditory adaptation process to the degraded new speech sound. The present findings suggest that the correct use of audio-visual bimodal training would facilitate the auditory rehabilitation process of patients with auditory prostheses such as a CI or ABI.

## 10.4 Summary

Visual information of audio-visual speech is known to complement degraded speech information in rehabilitation after implantation of a CI or ABI. In addition to this basic effect, audio-visual speech may also enhance the auditory adaptation process, with as little as a few hours audio-visual training.

**Acknowledgements** This study was supported by a grant-in-aid from the Japanese Ministry of Education, Culture, Sports, Science and Technology for fiscal 2007–2009 (Grant-in-Aid for Scientific Research (C) 19591954). Most contents of the present proceedings are based on our previous publications [9, 10].



## References

1. Fu QJ, Nogaki G, Galvin 3rd JJ. Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. J Assoc Res Otolaryngol. 2005;6:180–9.
2. 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;40:1562–9.
3. Halgren E, Dale AM, Sereno MI, Tootell RB, Marinkovic K, Rosen BR. Location of human face-selective cortex with respect to retinotopic areas. Hum Brain Mapp. 1999;7:29–37.
4. Haxby JV, Ungerleider LG, Clark VP, Schouten JL, Hoffman EA, Martin A. The effect of face inversion on activity in human neural systems for face and object perception. Neuron. 1999;22:189–99.
5. Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends Cogn Sci. 2000;4:223–33.
6. Hoffman EA, Haxby JV. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci. 2000;3:80–4.
7. Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci U S A. 1999;96:9379–84.
8. Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 1977;17:4302–11.
9. Kawase T, Yamaguchi K, Ogawa T, Suzuki K, Suzuki M, Itoh M, et al. Recruitment of fusiform face area associated with listening to degraded speech sounds in auditory-visual speech perception: a PET study. Neurosci Lett. 2005;382:254–8.

10. Kawase T, Sakamoto S, Hori Y, Maki A, Suzuki Y, Kobayashi T. Bimodal audio-visual training enhances auditory adaptation process. Neuroreport. 2009;20:1231–4.
11. McCarthy G, Puce A, Gore JC, Allison T. Face-specific processing in the human fusiform gyrus. J Cogn Neurosci. 1977;9:605–10.
12. Rosen SM, Fourcin AJ, Moore BC. Voice pitch as an aid to lipreading. Nature. 1981;291:150–2.
13. Sekiyama K, Kanno I, Miura S, Sugita Y. Auditory-visual speech perception examined by fMRI and PET. Neurosci Res. 2003;47:277–87.
14. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995;270:303–4.
15. Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J Acoust Soc Am. 1954;26:212–5.