

Toward Understanding Mechanisms of Sensorimotor Processing in Speech Production

Takemi Mochida[†] and Hiroaki Gomi

Abstract

We report on the online mechanisms of sensory feedback-based speech motor control. The neural processing mechanisms governing human speech production skills need to be understood in order to design a user-friendly and easy-to-use remote speech communication system. We elucidated the contributions of somatosensory and auditory feedback through experimental observation of involuntary responses induced by perturbed sensory feedback information during speech production tasks. The experimental results provide evidence that both the somatosensory feedback and auditory feedback associated with self speech production strongly affect the temporal dynamics of articulatory movements.

1. Introduction

In humans, speech production is a highly skilled aspect of muscular control and is gradually developed as a child matures. During the development, several different sorts of sensory feedback information play important roles in monitoring how well the action is being organized while various types of phonemes are produced. The sources for monitoring articulatory movements consist of cutaneous/somatosensory information related to the status of respiratory, laryngeal, velopharyngeal, and articulatory subsystems and auditory information representing the characteristics of the associated acoustic output.

1.1 Cutaneous/somatosensory feedback in speech production

The role of cutaneous and/or somatosensory feedback for speech motor control has been investigated in a series of studies [1]–[3] on compensatory articulatory movements of the upper lip induced by a perturbation of the jaw or lower lip during the production of the bilabial explosive consonants (/p/ and /b/). The

compensatory movements act effectively to achieve the intended acoustic sounds against unpredictable perturbation. One might speculate that an active compensatory mechanism recruited by somatosensory feedback contributes to the generation of these compensatory movements because the corresponding electromyographic (EMG) activity of the primary upper-lip muscle (orbicularis oris superior (OOS)) increased. However, the time delay due to nerve conduction and mechanochemical dynamics might present a problem for explanations of the rapid regulation of fast speech movements by sensorimotor coordination.

We have found that during the production of bilabial fricative consonant / Φ /, which requires precise control of the aperture between the upper and lower lips, the upper lip shifts downward rapidly in response to a sudden jaw-lowering perturbation to maintain the intact labial aperture [4]. Although the initial phase of the upper lip shift is generated by the mechanical linkage of perioral dynamics [4], the later phase will be partly regulated by the reactive muscle response. Actually, the upper lip muscle activity started to increase 48.25 ± 1.2 ms after the jaw perturbation. This latency is longer than that in a perioral reflex (14–17 ms [5]) that is mediated within the brainstem

[†] NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan

alone and shorter than the jaw's voluntary reaction time after a stimulus is perceived (150 ± 13 ms [6]). Considering these findings, it can be postulated that cortical processing is involved in this reflexive compensatory adjustment of speech articulation, as examined in the long latency stretch reflex [7]. We assessed the effect of transcranial magnetic stimulation (TMS) over the motor cortex on the reflexive compensatory adjustments of speech articulation [8]. Note that, in some cases, TMS can transiently disrupt or suspend cortical neural processing [9], [10], but in other cases it can enhance cortical neural excitability [7]. Our experimental results are presented in 2.1.

1.2 Auditory feedback in speech production

With regard to the role of auditory feedback for speech motor control, it is well known that speaking with exposure to a delayed auditory feedback (DAF) leads to various types of speech disfluencies, e.g., increased articulatory error, lengthened duration, augmented volume, and increased fundamental frequency [11]–[15]. Such disfluencies may occur as a result of several different types of voluntary and involuntary responses to DAF. The Lombard effect (or Lombard reflex) is well cited as an example of auditory-induced automatic motor response to a change in background noise level, where speakers involuntarily increase their vocal intensity as the noise level increases [16], [17]. Although these sorts of reflexive mechanisms can be considered as potential sources for such DAF-influenced speech disfluencies, the precise mechanisms of how the delayed auditory input of self-produced speech can adversely affect the speech motor control has not been fully elucidated yet.

Various studies using auditory feedback alteration have suggested that acoustic information is critical for learning and maintaining vowel production [18], [19] and voice pitch control [20], [21]. Evidence has also been obtained from humans and non-human primates showing that neural activity in the auditory cortex is modulated by self-produced vocalization [22]–[25]. However, there is an ongoing debate about whether such neural mechanisms also help ensure stability in rapid and complex speech motor control [26], [27]. Auditory feedback may serve as an immediate source for the dynamic control of speech articulation, analogous to the rapid adjustment of labial constriction based on cutaneous and/or somatosensory information. We examined the online control mechanism for articulatory lip movement during the repetition of bilabial plosives /pa/ by suddenly shift-

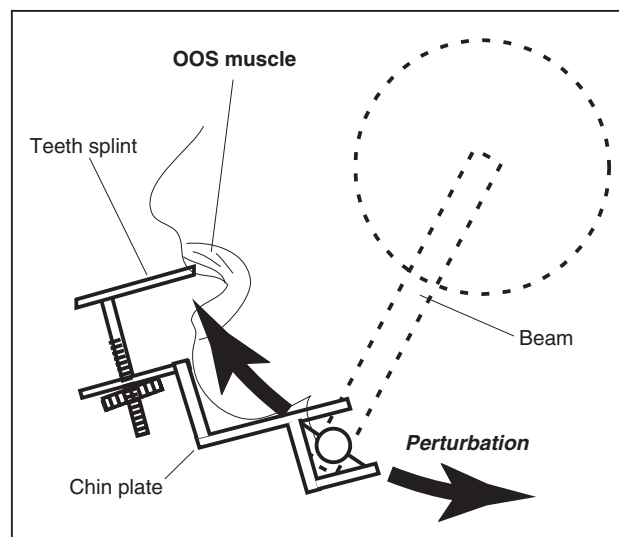


Fig. 1. Configuration of jaw perturbation system.

ing the auditory feedback timing in the ahead-of-time or delayed direction and/or by replacing the feedback syllable by other syllables [28]. Our experimental results are presented in 2.2.

2. Mechanisms of sensory feedback control: perturbation studies

2.1 Involvement of the motor cortex in reflexive speech motor coordination

We examined the facilitatory effect of TMS on the reflexive compensatory response to jaw perturbation during the production of bilabial fricative consonant / Φ /. The subject's jaw was held in a jaw perturbation system by clamping it between a chin plate and a custom-built splint that was attached to the teeth (**Fig. 1**). Note that this apparatus resulted in little disruption of normal speech and could apply a slight force in the jaw opening/closing direction. To avoid anticipation, a step-wise jaw-opening perturbation (3.0 N) was applied in 20% of the trials. Each session (100 trials) included the following conditions: PT (perturbation with TMS, 10 trials), PN (perturbation alone, 10 trials), and NT (TMS alone, 10 trials), and control (70 trials). There were three subjects denoted A–C. The jaw perturbation elicited a quick downward shift of the upper lip, accompanied by a muscle EMG response (black line in **Fig. 2(a)**) that served to maintain the labial aperture for producing / Φ /. The question here is what neural mechanism generates the EMG response. We compared EMG response

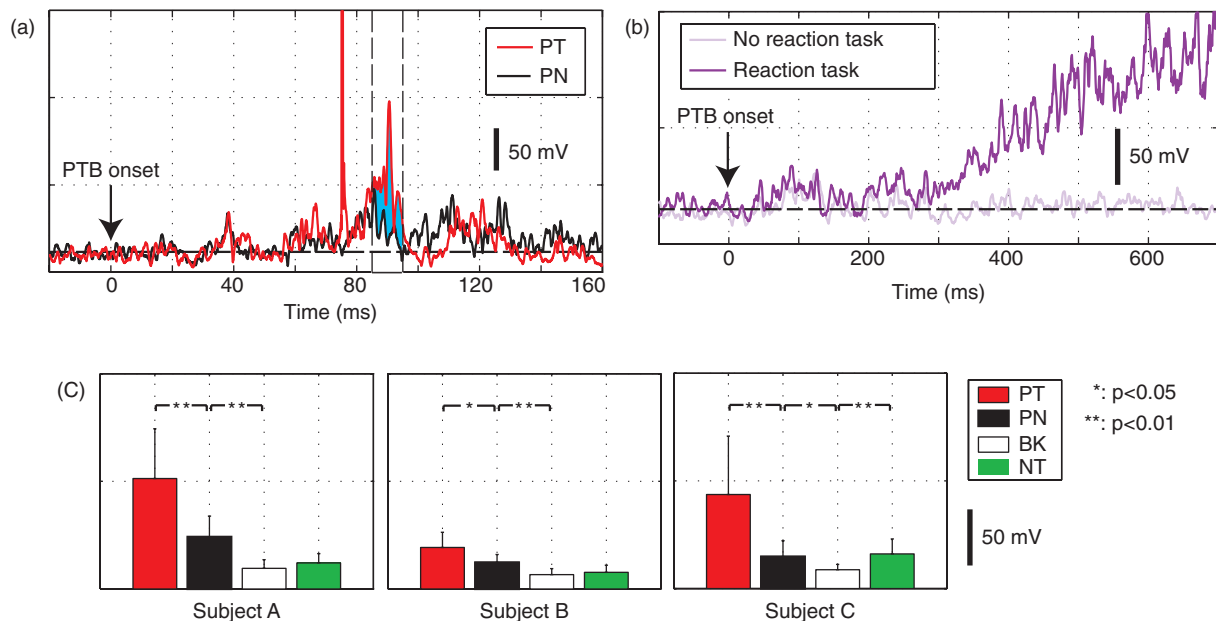


Fig. 2. (a) Typical reflexive compensatory response with or without TMS (EMGs of OOS), (b) EMGs of OOS in reaction or non-reaction task, and (c) comparison of response amplitudes.

latencies caused by the jaw perturbation with those involved in the voluntary reaction. A typical EMG response for the OOS in the reaction task is shown in **Fig. 2(b)**. The voluntary response started at around 300 ms after the perturbation. The mean reaction time for the three subjects (315.7 ± 98.4 SD ms; SD: standard deviation) was obviously longer than the latency of the reflexive compensatory response (48.25 ± 1.2 ms [4]), suggesting that the short-latency (< 100 ms) compensatory response was generated involuntarily.

We applied TMS in order to examine the involvement of the motor cortex in generating the reflexive compensatory response. We expected a TMS over the motor cortex to enhance the EMG activity of the response if the lip region of the motor cortex is involved in the reflexive compensatory response of the upper lip. The typical EMG pattern observed when TMS was applied during jaw perturbation (PT) is depicted by the red line in Fig. 2(a). The first sharp peak 75 ms after perturbation onset was an artifact induced by current spread due to TMS. Compared with the response without TMS (PN), an increase in EMG activity started 10 ms after TMS onset and continued for roughly 10 ms (shaded area). To quantify the amplitude of muscle response, the rectified EMG signal during a 10-ms window (10–20 ms after TMS onset) was temporally averaged and pooled in each condition (PT, PN, and NT). The background muscle

activity level (BK) was quantified by temporally averaging the rectified EMG signals for 10–20 ms prior to the perturbation (or stimulus) onset. The response amplitudes in all cases (PT, PN, BK, and NT) are summarized in **Fig. 2(c)**. TMS consistently enhanced the reflexive compensatory response in all subjects, as shown by the difference between the PT and PN cases (statistically significant), whereas there was no significant enhancement of muscle activity in NT compared with BK for subjects A and B. In the NT case for subject C, the muscle activity was slightly enhanced. The enhanced EMG activity in PT, however, was considerably higher than that in NT (difference between (PT minus PN) and (NT minus BK)), suggesting that the facilitatory effect was the primary determinant of the enhanced EMG activity in PT. In summary, these facilitations suggested that the cortical pathway contributes significantly to the production of the reflexive compensatory response.

2.2 Auditory-induced rapid change in articulatory lip movement

We evaluated changes in articulatory lip movement that occurred when a sudden alteration in auditory feedback timing and context was introduced while a subject was speaking the plosive-initial syllable /pa/ repetitively. A schematic diagram of the auditory feedback alteration system is shown in **Fig. 3**. The

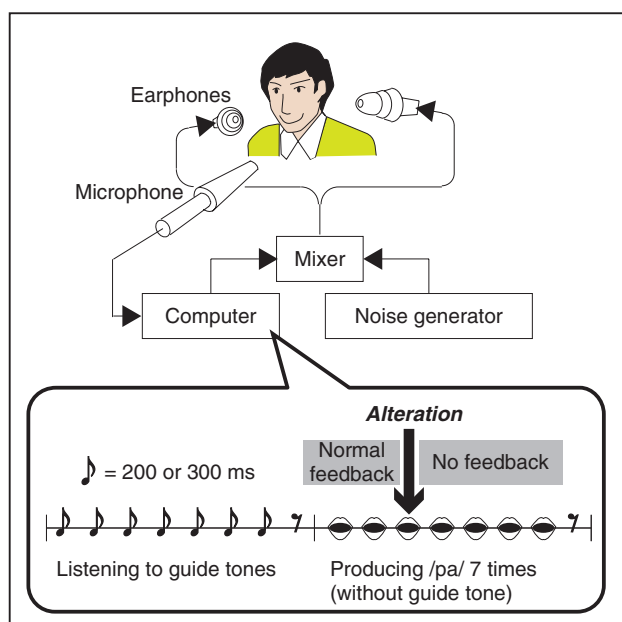


Fig. 3. Configuration of auditory feedback alteration system.

speech sounds produced by the subject were processed by a custom program running on a computer designed to alter the input speech signals. The altered signals were mixed with background noise and fed back to the subject's ears via earphones. Background noise can prevent subjects from hearing their own speech sounds while they are speaking. The subjects were asked to produce an isolated syllable /pa/ seven times while maintaining a constant speech rate. For each trial, the auditory feedback corresponding to the third repetition of /pa/ was altered by shifting the timing and/or replacing the type of syllable, while the subsequent feedback was omitted. Pre-recorded sounds /pa/, / Φ a/, and /pi/, spoken by the same subject, were used for the syllable replacement. The timing shift was -150, -100, -50, 0, +50, +100, or +150 ms from the third repetition onset, which was predicted from the interval between the onset of the first and second syllables in each trial. The three-dimensional motion of markers placed on the upper and lower lips was measured with an optical motion capture system, from which the aperture between the upper and lower lips (labial distance (LD)) was obtained. Typical LD trajectory data obtained during the production of /pa/ at a speech rate of 300 ms per syllable is shown in Fig. 4. The red curve in each panel shows the mean LD trajectory for five trials over the test blocks. The mean trajectory for ten trials

in the control (normal feedback condition) block is shown by a black curve. The LD trajectories in all conditions were temporally aligned at the predicted third syllable onset by referring to the simultaneously recorded acoustic signals. The auditory feedback conditions shown from the top to bottom panels were as follows: pre-recorded /pa/ was presented at -150, -100, -50, 0, 50, 100, and 150 ms from the predicted third syllable onset. The solid vertical line in each panel indicates the onset timing of the auditory stimulus, while the dotted vertical line indicates the predicted third syllable onset. By comparing the two trajectories in each panel, we found that the mouth opening movement subsequent to the auditory stimulus onset quickened by the -50-ms stimulus presentation. While a similar hasty movement was also observed for the -150- and -100-ms conditions, the effect seemed to be weaker. The deviation between the trajectories under each of the delayed feedback (50, 100, 150 ms) and control (0 ms) conditions was much smaller. Similar results were obtained for all ten subjects. The lags ($N = 10$) corresponding to the maximum cross-correlation between LD trajectories under the altered and control conditions within the post-stimulus period (dark shaded areas in Fig. 4), obtained by subtracting those within the pre-stimulus period (light shaded areas in Fig. 4) are shown in Fig. 5. Each pre- and post-stimulus period corresponds to a single cycle of lip closing/opening movement. A negative lag value reflects an ahead-of-time shift of the movement compared with the control. The condition indicated as *normal* refers to a comparison of the normal feedback trials during the test blocks and those in the control block, which reflects the variance in each subject's baseline speech rate throughout the experiment. The statistical significance of the difference from the normal condition ($p < 0.05$) was evaluated with a two-sided paired t-test (no. of degrees of freedom = 9 for all comparisons, with the Bonferroni adjustment). An ahead-of-time shift in the movement was found only when the auditory feedback of /pa/ preceded the real syllable production by 50 ms. An excessively early manipulation (-150 and -100 ms) of the auditory feedback did not significantly affect the movement. The delayed feedback (+50, 100, and 150 ms) also produced no significant change. Syllables that were not identical to those of the speech task (/ Φ a/ and /pi/) had no significant effect even when they were fed back 50 ms prior to the real syllable production. These results indicate that the ahead-of-time and delayed auditory feedback affected the articulatory lip movement in a time-

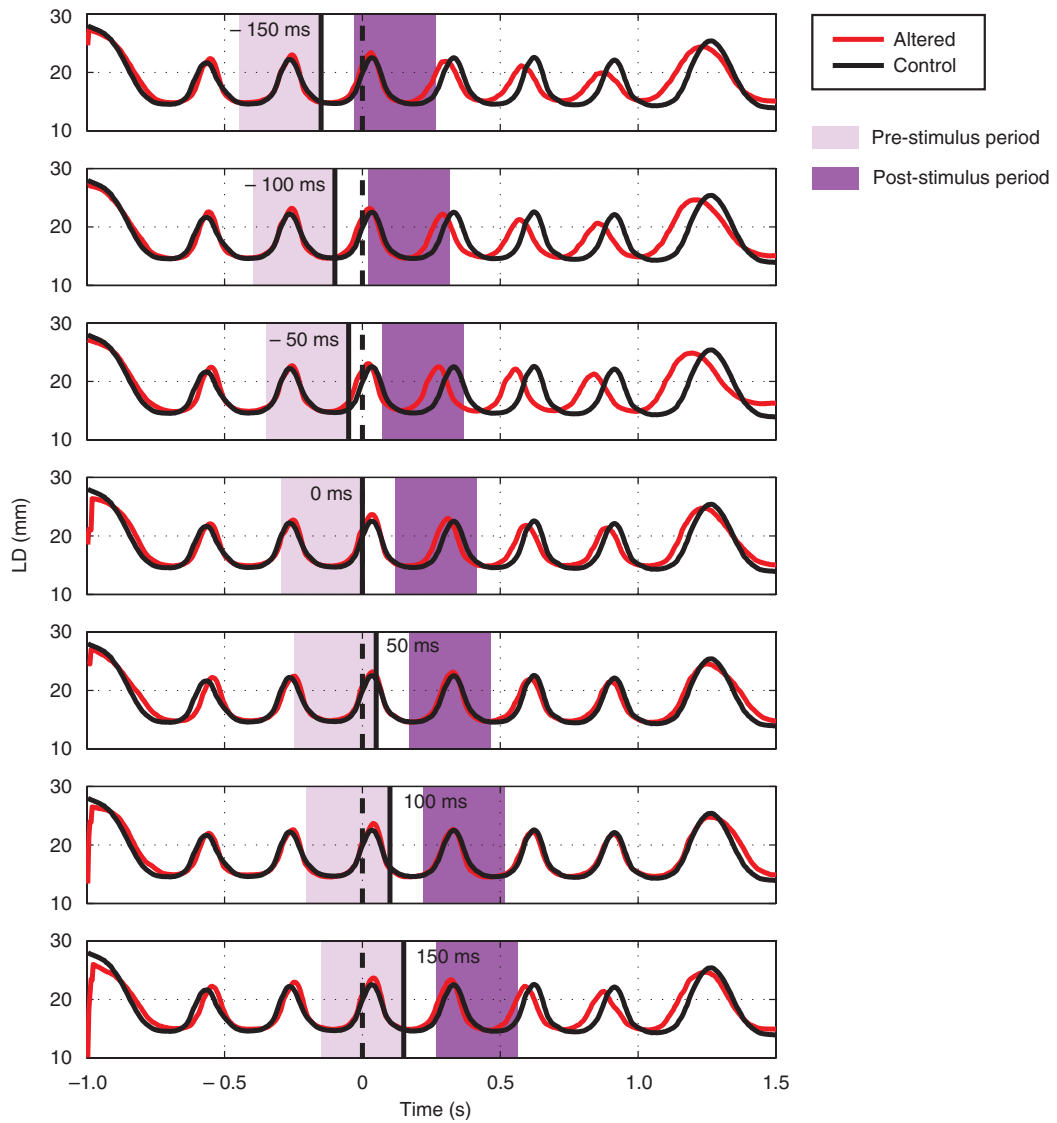


Fig. 4. Typical change in LD trajectories induced by time-shifted auditory feedback.

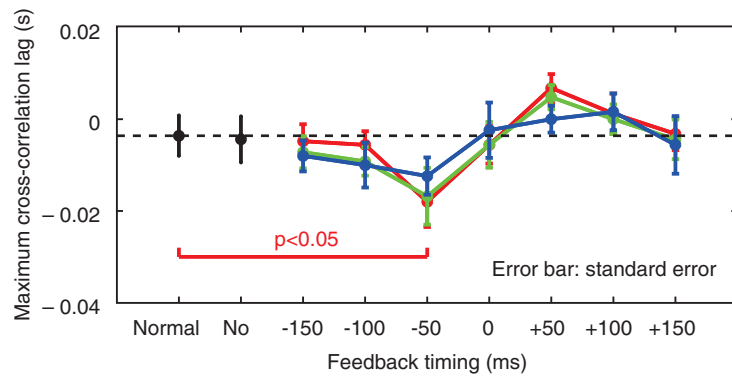


Fig. 5. Comparison of auditory-induced change in LD trajectories.

asymmetric and context-specific manner during repetitive syllable production. These findings suggest the existence of a compensatory mechanism to maintain a constant speech rate by detecting errors between the internally predicted and actually provided auditory information associated with self movement. The timing- and context-dependent effects of feedback alteration suggest that the sensory error detection works in a temporally asymmetric window where acoustic features of the syllable to be produced may be coded.

3. Conclusions

We investigated the contribution of sensory feedback to speech motor control by observing the involuntary response induced by perturbed somatosensory and auditory information. TMS to the cortex was demonstrated to have facilitatory effects on the reflexive compensatory response in lip muscles during labial speech production, and this led us to suggest that its generation involves the primary motor cortex. High-level computation in the cortex would greatly contribute to the organization of complex sensorimotor coordination among articulatory organs in order to achieve robustness in speech tasks. The articulatory lip movement quickened immediately when the auditory feedback preceded the expected timing by 50 ms. Such an articulatory change was not observed when the feedback was presented more than 50 ms earlier or later than the actual timing or when the feedback syllable was replaced by another syllable. These results suggest that errors between the internally predicted and actually provided auditory information detected in a temporally asymmetric window contribute to the compensation for the inter-articulatory timing in the syllable repetition task. Our studies provide evidence that the temporal dynamics of articulatory jaw and lip movements must be correctly maintained with both somatosensory and auditory feedback resulting from self speech production.

References

- [1] J. W. Folkins and J. H. Abbs, "Lip and Jaw Motor Control during Speech: Responses to Resistive Loading of the Jaw," *J. of Speech and Hearing Research*, Vol. 18, No. 1, pp. 207–219, 1975.
- [2] J. A. Kelso, B. Tuller, E. Vatikiotis-Bateson, and C. A. Fowler, "Functionally Specific Articulatory Cooperation Following Jaw Perturbations during Speech: Evidence for Coordinative Structures," *J. Exp. Psychol. Hum Percept Perform*, Vol. 10, No. 6, pp. 812–832, 1984.
- [3] S. Shaiman, "Kinematic and Electromyographic Responses to Perturbation of the Jaw," *J. Acoust. Soc. Am*, Vol. 86, No. 1, pp. 78–88, 1989.
- [4] H. Gomi, M. Honda, T. Ito, and E. Murano, "Compensatory Articulation during Bilabial Fricative Production by Regulating Muscle Stiffness," *J. of Phonetics*, Vol. 30, No. 3, pp. 261–279, 2002.
- [5] C. M. Weber and A. Smith, "Reflex Responses in Human Jaw, Lip, and Tongue Muscles Elicited by Mechanical Stimulation," *J. of Speech and Hearing Research*, Vol. 30, No. 1, pp. 70–79, 1987.
- [6] F. A. Ottenhoff, A. van der Bilt, H. W. van der Glas, and F. Bosman, "Peripherally Induced and Anticipating Elevator Muscle Activity during Simulated Chewing in Humans," *J. of Neurophysiol*, Vol. 67, No. 1, pp. 75–83, 1992.
- [7] B. L. Day, H. Riescher, A. Struppler, J. C. Rothwell, and C. D. Marsden, "Changes in the Response to Magnetic and Electrical Stimulation of the Motor Cortex Following Muscle Stretch in Man," *J. of Physiology*, Vol. 433, pp. 41–57, 1991.
- [8] T. Ito, T. Kimura, and H. Gomi, "The Motor Cortex is Involved in Reflexive Compensatory Adjustment of Speech Articulation," *NeuroReport*, Vol. 16, No. 16, pp. 1791–1794, 2005.
- [9] B. L. Day, J. C. Rothwell, P. D. Thompson, A. Maertens de Noordhout, K. Nakashima, K. Shannon, and C. D. Marsden, "Delay in the Execution of Voluntary Movement by Electrical or Magnetic Brain Stimulation in Intact Man—Evidence for the Storage of Motor Programs in the Brain," *Brain*, Oxford Univ. Press, UK, Vol. 112, No. 3, pp. 649–663, 1989.
- [10] T. Kimura, P. Haggard, and H. Gomi, "Transcranial Magnetic Stimulation over Sensorimotor Cortex Disrupts Anticipatory Reflex Gain Modulation for Skilled Action," *J. of Neurosci*, Vol. 26, No. 36, pp. 9272–9281, 2006.
- [11] B. S. Lee, "Effects of Delayed Speech Feedback," *J. of the Acoustical Society of America*, Vol. 22, No. 6, pp. 824–826, 1950.
- [12] W. R. Tiffany and C. N. Hanley, "Delayed Speech Feedback as a Test for Auditory Malingerer," *Science*, Vol. 115, No. 2977, pp. 59–60, 1952.
- [13] G. Fairbanks, "Selective Vocal Effects of Delayed Auditory Feedback," *J. of Speech and Hearing Disorders*, Vol. 20, No. 4, pp. 333–346, 1955.
- [14] S. Zanini, A. Clarici, F. Fabbro, and A. Bava, "Speaking Speed Effects on Delayed Auditory Feedback Disruption of Speech Fluency," *Percept Mot Skills*, Vol. 89, No. 3, Part 2, pp. 1095–1109, 1999.
- [15] A. Stuart, J. Kalinowski, M. P. Rastatter, and K. Lynch, "Effect of Delayed Auditory Feedback on Normal Speakers at Two Speech Rates," *J. of the Acoustical Society of America*, Vol. 111, No. 5, pp. 2237–2241, 2002.
- [16] E. Lombard, "Le Signe de L'elevation de la Voix" (English translation: "The Sign of Voice Rise"), *Annales des Maladies de L'Oreille et du Larynx*, Vol. 37, No. 2, pp. 101–119, 1911.
- [17] M. Garnier, N. Henrich, and D. Dubois, "Influence of Sound Immersion and Communicative Interaction on the Lombard Effect," *J. of Speech, Language, and Hearing Research*, Vol. 53, No. 3, pp. 588–608, 2010.
- [18] J. F. Houde and M. I. Jordan, "Sensorimotor Adaptation in Speech Production," *Science*, Vol. 279, No. 5354, pp. 1213–1216, 1998.
- [19] V. M. Villacorta, J. S. Perkell, and F. H. Guenther, "Sensorimotor Adaptation to Feedback Perturbations of Vowel Acoustics and Its Relation to Perception," *J. of the Acoustical Society of America*, Vol. 122, No. 4, pp. 2306–2319, 2007.
- [20] T. A. Burnett and C. R. Larson, "Early Pitch-shift Response is Active in Both Steady and Dynamic Voice Pitch Control," *J. of the Acoustical Society of America*, Vol. 112, No. 3, Part 1, pp. 1058–1063, 2002.
- [21] J. A. Jones and K. G. Munhall, "Remapping Auditory-motor Representations in Voice Production," *Current Biology*, Vol. 15, No. 19, pp. 1768–1772, 2005.
- [22] G. Curio, G. Neuloh, J. Numminen, V. Jousmäki, and R. Hari, "Speaking Modifies Voice-evoked Activity in the Human Auditory Cortex," *Hum Brain Mapp*, Vol. 9, No. 4, pp. 183–191, 2005.
- [23] S. J. Eliades and X. Wang, "Dynamics of Auditory-vocal Interaction in Monkey Auditory Cortex," *Cereb Cortex*, Vol. 15, No. 10, pp. 1510–1523, 2005.
- [24] T. H. Heinks-Maldonado, D. H. Mathalon, M. Gray, and J. M. Ford, "Fine-tuning of Auditory Cortex during Speech Production," *Psychophysiology*, Vol. 42, No. 2, pp. 180–190, pp. 180–190, 2005.

- [25] J. F. Houde, S. S. Nagarajan, K. Sekihara, and M. M. Merzenich, "Modulation of the Auditory Cortex during Speech: an MEG Study," *J. of Cognitive Neuroscience*, Vol. 14, No. 8, pp. 1125–1138, 2002.
- [26] G. J. Borden, "An Interpretation of Research of Feedback Interruption in Speech," *Brain Lang*, Vol. 7, No. 3, pp. 307–319, 1979.
- [27] A. Postma, "Detection of Errors during Speech Production: a Review of Speech Monitoring Models," *Cognition*, Vol. 77, No. 2, pp. 97–132, 2000.
- [28] T. Mochida, H. Gomi, and M. Kashino, "Rapid Change in Articulatory Lip Movement Induced by Preceding Auditory Feedback during Production of Bilabial Plosives," *PLoS One*, Vol. 5, No. 11, p. e13866, 2010.



Takemi Mochida

Research Scientist, Human and Information Science Laboratory, NTT Communication Science Laboratories.

He received the B.E. and M.E. degrees from Waseda University, Tokyo, in 1992 and 1994, respectively, and the Ph.D. degree from Future University-Hakodate in 2011. He joined NTT in 1994 and has been engaged in research on speech science. He was involved in JST-CREST projects from 1998 to 2003. His research interests include the sensorimotor mechanisms, articulatory-acoustic properties, and signal processing of speech. He is a member of the Society for Neuroscience, the Acoustical Society of America, the Japan Neuroscience Society, the Acoustical Society of Japan, and the Institute of Electronics, Information and Communication Engineers.



Hiroaki Gomi

Distinguished Senior Research Scientist, Group Leader of Sensory and Motor Research Group, NTT Communication Science Laboratories

He received the B.E., M.E., and Ph.D. degrees in mechanical engineering from Waseda University, Tokyo, in 1986, 1988, and 1994, respectively. He was involved in biological motor control research at ATR (Advanced Telecommunication Research Labs., Kyoto) from 1989 to 1994, in two JST-CREST projects from 1996 to 2003, and in a JST-ERATO project from 2005 to 2010. He was an adjunct professor of the University of Tokyo Institute of Technology from 2000 to 2004. His current research interests include human sensorimotor control mechanisms and interactions between perception and action.
