Novel approach for understanding the neural mechanisms of auditory-motor control: Pitch regulation by finger force

Ryosuke O. Tachibana [a,*], Masuzo Yanagida [b], Hiroshi Riquimaroux [a]

[a] Graduate School of Life and Medial Sciences, Doshisha University, Kyotanabe, 610-0394, Japan
[b] Graduate School of Engineering, Doshisha University, Kyotanabe, 610-0394, Japan

ABSTRACT

Music performance and speech production require neural circuits to integrate auditory information and motor commands to achieve rapid and accurate control of sound properties. This article proposes a novel approach for investigating neural substrates related to audiomotor integration. An experiment examined the brain activities involved in sensorimotor integration in a simplified audiomotor task: pitch regulation using finger-pinching force. The brain activities of the participants were measured using functional magnetic resonance imaging (fMRI) while they were performing the task. Two additional tasks were performed: an auditory-only task in which subjects listened to sound stimuli without any motor action and a motor-only task where they applied their finger force to the sensor in the absence of auditory feedback. The fMRI results showed the brain activities related to the online pitch regulation in the dorsal premotor cortex (dPMC), planum temporale (PT), primary auditory cortex, and part of the midbrain. The involvement of dPMC and PT was consistent with findings in previous studies on other audiomotor systems, implying that these regions appeared to be important for connecting the auditory feedback to motor actions.

© 2010 Elsevier Ireland Ltd. All rights reserved.

Music performance and speech production require rapid and accurate online control of sound properties. Successful achievement of this goal requires that motor responses reflect the information received via auditory feedback. Artificial delays in auditory feedback can produce stuttering in normal talkers [13] and interrupt successful performances of musicians [18,19]. Electrophysiological study focusing on a pianist's keystroke has suggested that the event-related potential elicited by errors in auditory feedback was increased when the pianist actually played piano [14]. These examples demonstrate the existence of neural circuits tightly connecting auditory perception with motor actions.

Neural substrates related to sensorimotor integration have been studied primarily with respect to the visual modality, using various experimental paradigms, such as reaching, grasping, ocular movements, and manipulation of a joystick or computer mouse. These studies have demonstrated that the posterior parietal cortex (PPC) and the premotor cortex (PMC) play roles in the transformation of visual cues into appropriate motor commands [1,8,9].

On the other hand, the neural substrates related to audiomotor integration have yet to be clearly identified. Neuroimaging studies using PET and fMRI technologies to examine audiomotor integration have focused on the online adjustment of vocalization.

A simple singing task, in which a prolonged vowel was vocalized at a constant pitch, activated the supplementary motor area, the anterior cingulate cortex, the precentral gyrus, the anterior insula, Heschl's gyrus, a posterior part of the superior temporal gyrus (STG), and parts of the cerebellum to a greater extent than did a simple listening task [17]. The delay and pitch shift in the auditory feedback during speech production increased activities in the STG, suggesting enhanced monitoring of one's own voice [5,7,15]. However, the shift in pitch reflected in the auditory feedback provided during simple singing activated a greater number of regions, including the PMC, the anterior insula, the intraparietal sulcus, and the supramarginal gyrus [22,25]. Additionally, many studies have reported the activation of the anterior insula during vocalization tasks [17,22,25]. Various regions have been proposed as candidates for the neural substrate of audiomotor integration, but no general agreement has emerged with respect to this issue.

The enigmatic character of neural substrates involved in audiomotor integration may be attributable to several limitations in vocalization experiments. First, it has been difficult to identify the function served by activated regions, because vocal control rests on a complex system including multiple neural pathways from brainstem nuclei to cerebral cortices. Second, vocalization frequently produces head movements that cause artifact noises to appear in MRI data. We also do not know whether the brain regions previously identified as candidates are vocal-specific or are more generally involved in audiomotor integration because

similar activities have not been confirmed with respect to non-vocal effectors.

In this study, we used fMRI to record brain activities in a simplified audiomotor situation in which the subjects were asked to manipulate tone pitch with their fingers. The fundamental frequency (F_0) of the generated tone changed in proportion to the force with which the subject gripped a designated object in real-time. The finger-grip task is a valuable tool for investigating sensorimotor integration [21,23], and the control system for the finger force is appear to be simpler than that for vocalization, having advantage on minimization of the head movements during fMRI scanning. This novel approach would also provide us insights for discrimination between effector-independent and effector-specific regions.

The purpose of this study was to demonstrate the usefulness of our novel approach in investigations of audiomotor integration such as those examining the brain activities involved in pitch regulation by finger force. Moreover, we assumed that the brain region involved in pitch regulation might be more activated if the amount of error correction was increased by pitch perturbation. Thus, we applied the pitch perturbation to the auditory feedback without evoking subjects' notice by slowly changing the relationships between force and pitch during the pitch regulation task.

Twelve subjects (7 males and 5 females, age between 21 and 27 years) participated in this experiment. They had musical experience as amateurs (e.g., playing the piano, violin, trumpet, for an average of 9.1 ± 5.4 years). No participant had absolute pitch. All subjects were right-handed and had normal hearing level (125–8000 Hz in octave steps, <15 dB HL). The study was approved

by the safety committee of the ATR Institute International and the ethics board of Doshisha University. Written informed consent was obtained from each subject before participation.

The force of each finger grip was measured isometrically with a strain-gauge-type force sensor (KEITEC System; Fig. 1A). The subject used the tips of the thumb and index finger of the right hand (i.e., precision grip) to grasp a sensor with flat parallel contact surfaces spaced 18 mm apart. The feedback tone of F_0 presented to each participant varied according to the force measured. In this article, the F_0 of all sounds is described in cents, logarithmically converted from Hertz using the equation: $\text{cents} = 1200 \times \log_2 (f/f_n)$, where f is the F_0 of the feedback tone in Hz and f_n is the frequency of the arbitrarily chosen note C5 (525.25 Hz). The F_0 of the feedback tone varied proportionally as a function of the finger force according to: $f_c = P - 500 + S$, where f_c denotes the F_0 of the feedback tone [cents], P denotes the finger force [gf], and S denotes the F_0 shift in cents for varying force targets and perturbation conditions (Fig. 1B). Thus, force at 500 gf produced F_0 at 0 cents (525.25 Hz) when $S = 0$. The force change spanning +1 gf corresponds to the F_0 change spanning +1 cents, regardless of S. The body of the sensor consisted of plastic, aluminum, and titanium, and the contact surfaces were made of thin felt and hard rubber. The waveform of the feedback tone was a saw-tooth signal consisting of the lowest four harmonics (-6 dB/oct, 20 ms rise/decay). The real-time generation of the feedback tone was achieved by a digital signal processor (s-BOX, MTT).

Subjects were asked to perform three types of experimental tasks: audiomotor (AM), motor-only (M), and auditory-only (A).
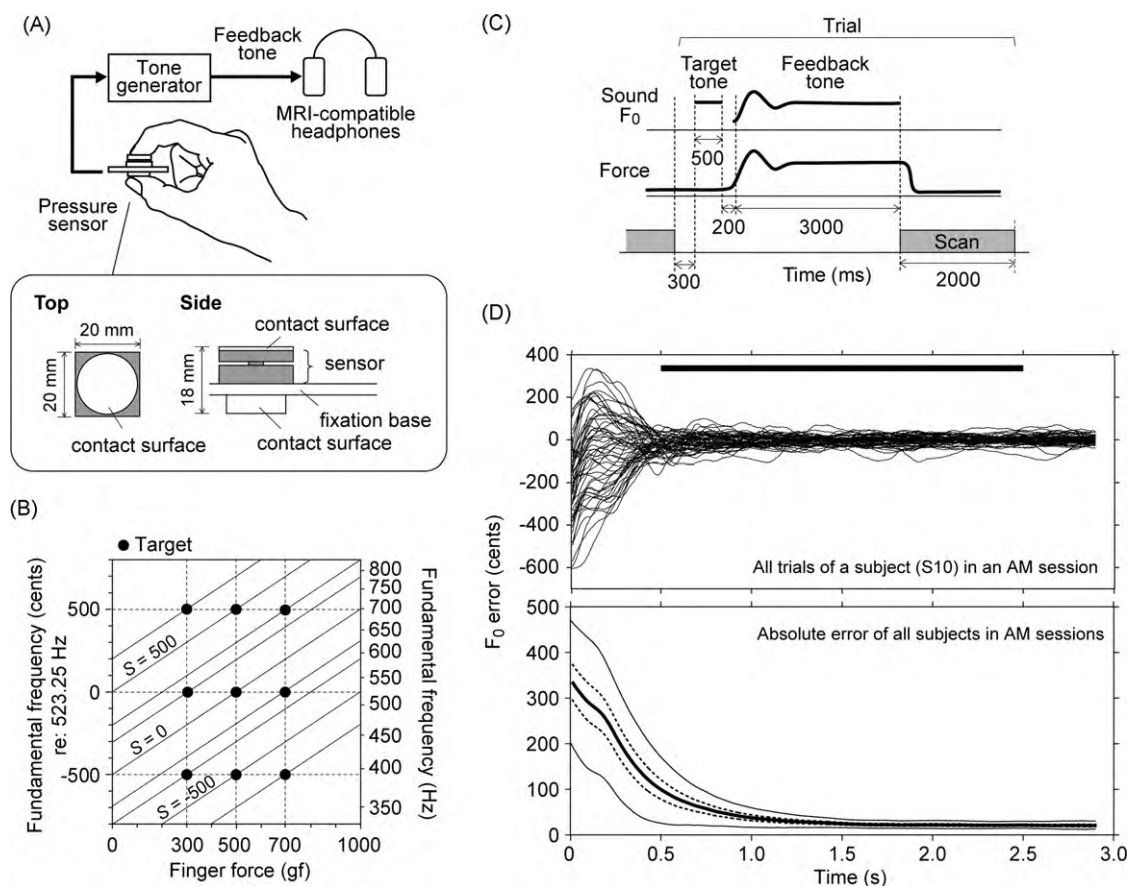

Fig. 1. (A) Experimental arrangement and details of the force sensor. (B) Relationship between the finger force and the F_0 of the feedback tone. Black dots indicate the target conditions. The parallel oblique line represents the relationship under each target condition. S indicates the F_0 shifts in cents. (C) Illustration of trial procedure. Top, middle, and bottom parts indicate the F_0 of sound stimuli, the applied finger force, and the scanning sequence, respectively. (D) Examples of F_0 data. The time courses of F_0 errors committed by one subject (S10) in all trials during an AM session are shown in the upper panel. The bold black bar indicates the duration of the pitch-shift perturbation. The mean (bold line), the standard deviation (thin), and the standard error of the mean (dashed) of the absolute F_0 error obtained from all subjects are shown in the lower panel.

The M and A tasks served as baseline tasks to compare with the AM task. The AM task contained two conditions, with and without gradual F_0-shift perturbations.

In the AM task, the subjects were asked to regulate and match the pitch of the feedback tone to the pitch of the target tone as quickly and precisely as possible. At the beginning of a trial, the target tone was presented for 500 ms and the feedback tone was presented for 3000 ms after a 200 ms interval (Fig. 1C). The inter-trial interval was 6000 ms. The F_0 of the target tone (target F_0) was set at −500, 0, or 500 cents (G4, C5, or F5). The force required to reach the target F_0 (target force) was set at 300, 500, or 700 gf. Thus, nine target conditions (3 F_0s × 3 forces) were included in the experimental design (Fig. 1B). The target F_0 and target force varied among trials. The upper limit of target force (700 gf) was designed to be less than 20% of the mean maximum contraction of all subjects (3945.1 gf). Under the with-perturbation condition (AMp), the F_0 of the feedback tone gradually shifted at a rate of ±50 cents/s. The perturbation was presented between 500 and 2500 ms after the onset of the feedback tone, and the F_0 shift was either +100 or −100 cents. Under the without-perturbation condition (AMn), the F_0 was not perturbed. Training trials of approximately 90 min were performed for the AM task. A total of 540 trials (9 targets × 3 types of perturbation (none, up, down) × 20 trials) were presented in random order during training. Short breaks of several minutes were inserted on request. The entire training was conducted in 1 day within 4 days before the day of fMRI scanning. An additional 270 trials were presented to one subject (S6) because his/her control error had been unstable during the first half of the training. No subject was aware of the presence of perturbations during the experiment, but all were able to successfully maintain the F_0 of the feedback tone after the training. In the M task, one of three instruction sounds was presented at the beginning of each trial, and subjects were then asked to press the sensor with a finger force corresponding to the instruction in the absence of any auditory feedback. The instruction sound consisted of single or multiple short band-noise bursts (100–6000 Hz, 10 ms rise/decay). We instructed subjects that single noise burst of 300 ms duration, double 150 ms bursts with a 50 ms gap, and triple 100 ms bursts with two 50 ms gaps indicated the target forces of 300, 500, and 700 gf, respectively. Subjects were trained to perform the task correctly before fMRI scanning. During the training, a white noise was auditorily fed back to participants in real-time when the finger force was within the target force ±100 gf. Finally, all subjects learned to generate the correct force with no auditory feedback. In the A task, subjects were asked to just listen to two-tone bursts without any motor action. After presentation of the first tone, which was identical to the target tone in the AM task, the second tone, which was 3000 ms in duration and characterized by the same F_0 as the first, was presented following a 200 ms gap. Subjects were instructed to avoid physical motion and the imagining of any finger movement. The waveforms of the tone stimuli used in the AM and A tasks were identical to that of the feedback tone.

A sparse temporal imaging technique [4] was used to avoid the acoustic and electromagnetic interference produced by scanning. Two-second scans were repeated every 6 s, and experimental tasks were presented during the 4 s interval between successive scans that were devoid of scanning noise. Scanning sessions were conducted separately for AM, M, and A tasks. In the AM session, two or three successive trials under the AMp or AMn condition comprised one block, and the AMp and AMn blocks were presented alternately. Two baseline trials in which the subject rested but performed no action were inserted every eight AM trials. The perturbation shift-up and shift-down conditions were presented alternately between AMp blocks. In the M or A session, two baseline trials were inserted every nine M or A trials, respectively. In total, 54 trials were conducted under each of the conditions.

A 3-Tesla MRI system (Magnetom Trio, Siemens) was used. Echo-planar imaging (EPI) volumes consisting of 3 mm × 3 mm × 5 mm voxels were obtained within 2 s with a repetition time of 6 s (30 slices; 4 mm thickness with 1 mm gap; transverse; FOV: 192 mm × 192 mm; matrix size: 64 × 64; TE: 30 ms; FA: 90°). Structural images were scanned with a 1 mm × 1 mm × 1 mm resolution (192 1 mm thick slices; sagittal; FOV: 256; matrix size: 256 × 256; TI: 900 ms; TR: 2250 ms; TE: 3.06 ms; FA: 9°). Sound stimuli were presented in the scanner room to subjects via MRI-compatible headphones (Hitachi Advanced Systems) while their eyes were closed. The force sensor was tightly fixed on the bedside guide rail of the scanner to avoid electromagnetic induction.

For purposes of the analysis of behavioral data, the F_0 was calculated at 5 ms intervals during each trial. The F_0 error in each trial was defined as the average of the absolute differences in cents between the target and the feedback tones within 2 s from 500 to 2500 ms after initiation of the feedback tone (shown in Fig. 1D, as a black bar in the upper panel).

Functional imaging data were processed and analyzed using the SPM5 software (Wellcome Department of Cognitive Neurology, London, UK). The scanned images obtained from each subject were realigned, spatially normalized into the standard stereotaxic space, and smoothed by an 8 mm Gaussian kernel. The spatially pre-processed data were then statistically analyzed with the general linear model using a voxel-by-voxel approach [10] as follows. The regressor representing the on/off of trials was modeled for each type of task (i.e., AMn, AMp, M, and A). The duration of the trials for all types of tasks corresponded to the duration of the feedback tone under the AM condition. A subtractive contrast '(AMn + AMp) − (A + M)' was used to identify audiomotor-related activities. This contrast can eliminate irrelevant regions commonly activated by all tasks. Estimated contrast images derived from single-subject analyses were included in the group analysis using a one-sample t-test as a random-effect model ($p < 0.01$, corrected for a false discovery rate, FDR). Clusters for volume sizes less than 10 voxels were eliminated. Additionally, the differences in the brain activities within the audiomotor-related regions under the AMn and AMp conditions were examined using a two-sample t-test ($p < 0.05$, FDR corrected). Voxels whose values were significantly lower during tasks than during rest ($p < 0.001$, uncorrected) were eliminated from all analysis. The realignment parameters were included in the model as a regressor to remove potential artifacts related to head movement. The regressors were convolved with basis functions consisting of a canonical hemodynamic response and its time and dispersion derivatives. Low-frequency drifts were removed using a high-pass filter with a cut-off period of 128 s, and serial correlations among the scans in each session were removed with a first-order autoregressive model.

With respect to the behavioral results, the F_0 errors in both AMn and AMp trials decreased as a function of the number of trials. The average error in the last block of training (M ± SD: 23.4 ± 10.4 cents) was significantly lower than that in the first block (39.1 ± 17.9 cents; paired t-test, $p < 0.01$) when trials during training were equally separated into 20 blocks. No significant difference was observed between the F_0 error values in the last block of training (23.4 ± 10.4 cents) and those in fMRI scanning (28.7 ± 12.8 cents). The average finger force in the AM and M trials during fMRI scanning did not differ significantly (Wilcoxon's signed-rank test, $p > 0.05$).

The fMRI results showed brain activities related to audiomotor integration in the frontal and the temporal lobes, the cerebellar cortex, and the midbrain (Fig. 2A; Table 1). Activation peaks were found in the primary motor cortex (M1) and the dorsal premotor cortex (dPMC) of the left hemisphere, the pre-supplementary motor area (preSMA) on the interhemispheric portion, and the pri-
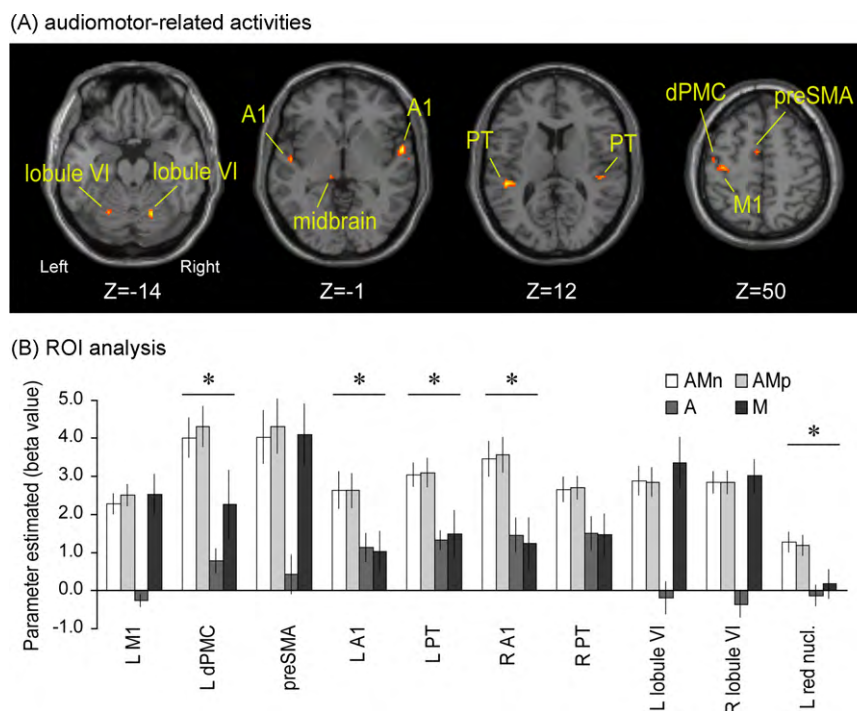
Fig. 2. (A) Activation map related to the audiomotor task. (B) Results of the region-of-interest analysis. The asterisks indicate regions in which the activations for both the AMn and AMp tasks were significantly higher than were those for the A and M tasks according to pair-wise comparisons. The error bar represents the standard error of the mean.

mary auditory cortex (A1) and the planum temporale area (PT) of the bilateral hemisphere. The hemispheric lobules VI in the bilateral cerebellar cortex and the left midbrain also showed activation peaks. We found no significant difference between the activations under the AMn and AMp conditions.

The activation height (estimated beta values) within spherical region-of-interest (ROI) centered at each of the activation peaks with a 5 mm radius were extracted (Fig. 2B). Pair-wise comparisons in each ROI showed that the activations for both the AMn and AMp tasks were significantly higher than were those for the A and M tasks in the bilateral A1, the left dPMC, the left PT, and the midbrain ($p < 0.05$, corrected by the Tukey–Kramer method). In other

Table 1

Representative activation peaks related to the audiomotor integration tested by the statistical contrast: $(AMn + AMp) - (A + M)$ ($p < 0.01$, FDR corrected; cluster size ≥ 10 voxels).

Region	Talairach coordinates			t value
	x	y	z	
Frontal lobe				
L: M1 (BA4)	−36	−15	50	9.18
L: dPMC (BA6)	−48	−5	52	7.44
M: preSMA (BA6)	0	1	55	14.81[*]
Temporal lobe				
L: A1 (BA22)	−50	−10	−1	8.87
L: PT (BA41)	−46	−32	13	11.53[*]
R: A1 (BA22)	59	−2	0	13.70[*]
R: PT (BA41)	48	−27	12	8.54
Cerebellum				
L: hemispheric lobule VI	−22	−63	−14	9.05
R: hemispheric lobule VI	20	−63	−15	13.19[*]
Midbrain				
L: red nucleus	−10	−25	−2	8.08

M1, primary motor cortex; dPMC, dorsal premotor cortex; preSMA, pre-supplementary motor area; A1, primary auditory cortex; PT, planum temporale; L, left; R, right; M, medial.

[*] Peaks that survived the threshold at $p < 0.05$ (FWE corrected).

words, activations in these regions would be related to the online pitch adjustment. No significant difference between the activities induced by the AM and M tasks emerged in the preSMA, the cerebellar cortex, and the M1, indicating that these regions were involved in motor control only. The auditory feedback in AM trial had a varying pitch while the pitch was not varied in A trial. It is possible to assume that the difference in pitch variation caused additional activities in the brain regions involved in basic auditory perception, such as A1.

Previous studies on vocalization have reported the similar activities including the premotor cortex and posterior part of STG [17,22,25]. The PT has been predicted to play a role in connecting auditorily processed information with other sensory and motor modalities [3,6,16,24]. Visuomotor studies have shown an association between the dPMC and motor-response selections conditioned by sensory cues [8]. Several studies have demonstrated a similar integrative function for the dPMC in audiomotor situations [2,12]. These findings suggest that both the PT and dPMC are very important for the online pitch adjustment, analogous to the PPC–PMC network for visuomotor control [1,8,9].

In this experiment, subjects had to keep the target pitch briefly in their mind in AM condition whereas they did not in A condition. This fact implies possibility that activations involved in AM task contained components related to the working memory. Interestingly, Koelsch et al. [11] reported that brain regions similar to the present study were involved in the auditory working memory, and suggested an involvement of vocal sensorimotor processes in the auditory working memory. The present data, however, did not provide critical accounts for relationship between working memory and auditory-motor interactive process. Further studies are needed on this issue.

Although many previous studies on vocal control have reported activities in the insular cortex, particularly in the anterior insula [17,22,25], we observed no significant activation in this area, suggesting that the activations in the insular cortex during vocalization tasks were related to such vocal-specific functions as motor con-

trol over vocal organs [20], rather than to the effector-independent function involving audiomotor integration.

In contrast to results from previous studies of audiovocal control using pitch perturbations [22,25], in the present study, the introduction of gradual perturbations did not elicit significant differences in brain activity. Previous studies found that additional brain activities compensated for a stepwise F_0 shift of 200 cents up/down in auditory feedback during vocalization. In this study, the subjects were not aware of the presence of perturbations because shifts were introduced gradually at the rate of ± 50 cents/s. The lack of subjective awareness of deliberately induced perturbations or an insufficient F_0 shift in the perturbation might have eliminated the additional brain activities observed in previous studies.

This study showed brain regions involved in the online pitch adjustment by using our novel task that required pitch regulation by finger-pinching force. Comparing results of the present study with those of previous studies on audiovocal and visuomotor control suggested that the dPMC and PT activations were parts of important components of the neural circuits connecting auditory feedback with motor actions.

Acknowledgements

References

[1] C.A. Buneo, M.R. Jarvis, A.P. Batista, R.A. Andersen, Direct visuomotor transformations for reaching, Nature 416 (2002) 632–636.

[2] J.L. Chen, R.J. Zatorre, V.B. Penhune, Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms, Neuroimage 32 (2006) 1771–1781.

[3] T.D. Griffiths, J.D. Warren, The planum temporale as a computational hub, Trends Neurosci. 25 (2002) 348–353.

[4] D.A. Hall, M.P. Haggard, M.A. Akeroyd, A.R. Palmer, A.Q. Summerfield, M.R. Elliott, E.M. Gurney, R.W. Bowtell, "Sparse" temporal sampling in auditory fMRI, Hum. Brain Mapp. 7 (1999) 213–223.

[5] Y. Hashimoto, K.L. Sakai, Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study, Hum. Brain Mapp. 20 (2003) 22–28.

[6] G. Hickok, D. Poeppel, The cortical organization of speech processing, Nat. Rev. Neurosci. 8 (2007) 393–402.

[7] S. Hirano, H. Kojima, Y. Naito, I. Honjo, Y. Kamoto, H. Okazawa, K. Ishizu, Y. Yonekura, Y. Nagahama, H. Fukuyama, J. Konishi, Cortical processing mechanism for vocalization with auditory verbal feedback, Neuroreport 8 (1997) 2379–2382.

[8] E. Hoshi, J. Tanji, Integration of target and body-part information in the premotor cortex when planning action, Nature 408 (2000) 466–470.

[9] P.B. Johnson, S. Ferraina, L. Bianchi, R. Caminiti, Cortical networks for visual reaching: physiological and anatomical organization of frontal and parietal lobe arm regions, Cereb. Cortex 6 (1996) 102–119.

[10] S. Kiebel, A. Holmes, The general linear model, in: R.S.J. Frackowiak (Ed.), Human Brain Function, Academic Press, 2004, pp. 725–760.

[11] S. Koelsch, K. Schulze, D. Sammler, T. Fritz, K. Müller, O. Gruber, Functional architecture of verbal and tonal working memory: an FMRI study, Hum. Brain Mapp. 30 (2009) 859–873.

[12] K. Kurata, T. Tsuji, S. Naraki, M. Seino, Y. Abe, Activation of the dorsal premotor cortex and pre-supplementary motor area of humans during an auditory conditional motor task, J. Neurophysiol. 84 (2000) 1667–1672.

[13] B.S. Lee, Effects of delayed speech feedback, J. Acoust. Soc. Am. 22 (1950) 639–640.

[14] C. Maidhof, N. Vavatzanidis, W. Prinz, M. Rieger, S. Koelsch, Processing expectancy violations during music performance and perception: an ERP study, J. Cogn. Neurosci. 22 (2010) 2401–2413.

[15] P.K. McGuire, D.A. Silbersweig, C.D. Frith, Functional neuroanatomy of verbal self-monitoring, Brain 119 (1996) 907–917.

[16] T. Paus, D.W. Perry, R.J. Zatorre, K.J. Worsley, A.C. Evans, Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges, Eur. J. Neurosci. 8 (1996) 2236–2246.

[17] D.W. Perry, R.J. Zatorre, M. Petrides, B. Alivisatos, E. Meyer, A.C. Evans, Localization of cerebral activity during simple singing, Neuroreport 10 (1999) 3979–3984.

[18] P.Q. Pfordresher, C. Palmer, Effects of hearing the past, present, or future during music performance, Percept. Psychophys. 68 (2006) 362–376.

[19] B.H. Repp, Effects of auditory feedback deprivation on expressive piano performance, Music Percept. 16 (1999) 409–438.

[20] A. Riecker, H. Ackermann, D. Wildgruber, G. Dogil, W. Grodd, Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum, Neuroreport 10 (2000) 1997–2000.

[21] A. Sterr, S. Shen, C. Kranczioch, A.J. Szameitat, W. Hou, B. Sorger, fMRI effects of task demand and feedback accuracy on grip force tracking, Neurosci. Lett. 458 (2009) 61–65.

[22] Toyomura, S. Koyama, T. Miyamaoto, A. Terao, T. Omori, H. Murohashi, S. Kuriki, Neural correlates of auditory feedback control in human, Neuroscience 146 (2007) 499–503.

[23] D.E. Vaillancourt, K.R. Thulborn, D.M. Corcos, Neural basis for the processes that underlie visually guided and internally guided force control in humans, J. Neurophysiol. 90 (2003) 3330–3340.

[24] J.E. Warren, R.J.S. Wise, J.D. Warren, Sounds do-able: auditory-motor transformations and the posterior temporal plane, Trends Neurosci. 28 (2005) 636–643.

[25] J.M. Zarate, R.J. Zatorre, Experience-dependent neural substrates involved in vocal pitch regulation during singing, Neuroimage 40 (2008) 1871–1887.