Experimental Brain Research (2019) 237:2197–2204
https://doi.org/10.1007/s00221-019-05584-1

RESEARCH ARTICLE

Precision of voicing perceptual identification is altered in association with voice-onset time production changes

Shunsuke Tamura[1] · Kazuhito Ito[2] · Nobuyuki Hirose[3] · Shuji Mori[3]

Received: 26 November 2018 / Accepted: 13 June 2019 / Published online: 19 June 2019

Abstract

There is ample evidence that motor learning changes the function of perceptual systems. Previous studies examining the interactions between speech production and perception have shown that the discrimination of phonetic contrasts characterized by the difference in articulatory place features is altered following their production changes caused by the perturbation of auditory feedback. The present study focused on a voiced–voiceless contrast in stop consonants, which is characterized by a temporal articulatory parameter, voice-onset time (VOT). In the experiment, we manipulated the participants' motor functions concerning VOT using a cross-categorical auditory feedback (CAF) paradigm (Mitsuya et al. in J Acoust Soc Am 135:2986–2994, 2014), in which a pre-recorded syllable sound starting with a voiced stop consonant (/da/) was fed back simultaneously with the participant's utterance of a voiceless stop consonant (/ta/), and vice versa. The VOT difference between /da/ and /ta/ productions was increased by the CAF, which is consistent with the result of Mitsuya's study. In addition, we conducted perceptual identification tasks of /da/-/ta/ continuum stimuli varying in VOT before and after the CAF task, and found that the identification function became sharper after as compared to before the CAF task. A significant positive correlation between such production and perception changes was also found. On the basis of these results, we consider that the change in motor function concerning VOT affected voiced–voiceless perceptual processing. The present study is the first to show the involvement of the speech production system in the perception of phonetic contrasts characterized by articulatory temporal features.

Keywords Voicing production · Voicing perception · Voice-onset time · Cross-categorical auditory feedback

Introduction

Mapping highly variable acoustic speech sounds to discrete phonetic categories is the most fundamental function of speech perception (Liberman et al. 1967). However, its underlying mechanism has not been clarified. There are hypotheses that the speech production system is closely involved with speech perception (Liberman and Mattingly 1985; Stevens and Halle, 1967), although several studies stressed the importance of auditory-based speech processing for phonetic perception (Holt et al. 2004; Stevens 1989). Specifically, Stevens and Halle suggested that the motor system assists phonetic perception by providing production-based constraints on the analysis of speech sounds. Liberman and Mattingly maintained that speech sounds are phonetically perceived by estimating the articulatory gestures producing them. The existence of such motor-based speech processing for phonetic perception was supported by a number of neurophysiological findings that the speech motor cortices located in the left hemisphere were activated when perceiving phonetic speech sounds (Chevillet et al. 2013; Evans and Davis 2015; Lee et al. 2012; Pulvermüller et al. 2006; Schomers and Pulvermüller 2016; Skipper et al. 2017; Wilson et al. 2004).

In several recent studies, it was suggested that both auditory-based processing and motor-based processing of speech sounds contribute to phonetic perception (Devlin and

✉ Shunsuke Tamura
tamuras@cog.inf.kyushu-u.ac.jp

[1] Department of Informatics, Graduate School of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Fukuoka 819-0395, Japan

[2] Department of Information Science, Faculty of Liberal Arts, Tohoku Gakuin University, 2-1-1 Tenjinzawa, Izumi-ku, Sendai, Miyagi 981-3193, Japan

[3] Department of Informatics, Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Fukuoka 819-0395, Japan

Aydelott 2009; Hickok et al. 2011; Schwartz et al. 2012). A current theory regarding speech processing in the brain suggests that speech sounds are cortically processed concurrently in both ventral (from the auditory cortex to the inferior frontal cortex) and dorsal streams (from the auditory cortex to the speech motor and the inferior frontal cortices via the inferior parietal lobe) in the left hemisphere and that the dorsal stream is involved in mediating the auditory and motor representations of speech sounds during speech production and perception, while the ventral stream is involved in processing speech sounds for comprehension on the basis of auditory information (Rauschecker and Scott 2009; Rauschecker 2011).

The involvement of the motor system with the perception system is supported by a number of behavioral findings showing that motor learning changes perceptual systems, in particular, visual and somatosensory systems, and their networks in the brain (Ostry and Gribble 2016). Concerning the auditory system, there are several studies examining the contribution of speech production systems to speech perception. Shiller et al. (2009) and Lametti et al. (2014) suggested that the modifications of the speech production system, caused by a transformed auditory feedback (TAF) task, altered phonetic perceptual processing. Shiller et al. (2009) found that the centroid frequency of a sibilant consonant /s/ production increased slightly during a TAF task in which a participant uttered a word beginning with /s/ and was presented in synchrony with the participant's utterance for which the centroid frequency had been slightly decreased toward /ʃ/ (the feedback sound was heard as /s/). In addition, they found that the perceptual boundary of /s/-/ʃ/ continuum stimuli varying in formant amplitude shifted in the direction of /ʃ/ after as compared to before the TAF task. Lametti et al. (2014) also found that the perceptual boundary of /ɛ/-/æ/ or /ɛ/-/ɪ/ continuum stimuli shifted in the direction of /ɛ/ along with a decrease or increase in the first formant frequency of a phoneme /ɛ/ production toward /æ/ or /ɪ/, caused by a TAF task in which participants uttered a word including a phoneme /ɛ/ and were presented in synchrony with their utterance sound for which the first formant had been slightly increased or decreased toward those of /ɪ/ or /æ/.

Patri et al. (2018) assumed that both auditory-based speech processing and the motor-based speech processing contribute to phonetic perception and discussed what functional changes, caused by the TAF task, drive the changes of phonetic production and perception by considering the experimental results of Lametti et al. (2014) in a Bayesian modeling framework. They suggested that the perturbation of auditory feedback updates the auditory characterizations of a produced phoneme, and that these updates alter phonetic production in the same direction as the acoustic manipulation of the auditory feedback because phonetic production is conducted towards a goal of auditory characterization of a produced phoneme. In addition, they suggested that the auditory–motor internal models are also updated by the perturbation of auditory feedback to reduce the mismatch between the auditory target of a produced phoneme and auditory feedback and that these updates affect phonetic production in the opposite direction of the acoustic manipulation of the auditory feedback as a number of previous studies have suggested (Houde and Jordan 1998, 2002; Jones and Munhall 2000, 2005; Villacorta et al. 2007). Concerning perception changes caused by the TAF, it was suggested that the updates of the auditory characterizations of the produced phoneme and the auditory–motor internal models altered auditory-based speech processing and the motor-based speech processing for phonetic perception, respectively.

Most previous studies examined the effect of the TAF task on articulatory place (acoustically spectral) features of speech to investigate the underlying mechanism of phonetic production and its connection to phonetic perception. On the other hand, few studies focused on articulatory temporal (acoustically temporal) features because real-time manipulation of acoustical temporal features is not easily implemented in the TAF task. Mitsuya et al. (2014) focused on an articulatory temporal parameter, voice-onset time (VOT), which is defined as the time interval between the onsets of consonant release and periodic vocal cord vibrations (Lisker and Abramson 1964), and examined whether the perturbation of auditory feedback affected VOT productions of voiced and voiceless stop consonants. They conducted a cross-categorical auditory feedback (CAF) task in which a pre-recorded word sound starting with a voiced stop consonant (/d/) was fed back almost simultaneously with the participant's utterance of a word starting with /t/, and vice versa, and found that the CAF increased the VOT difference between /d/ and /t/ productions. They focused on the difference in manipulation of auditory feedback between the TAF and CAF tasks and suggested that CAF-induced VOT production changes were not due to feedback control to reduce the auditory feedback error, which corresponds to an update of the auditory–motor internal models (Patri et al. 2018), but rather due to feedforward control to maintain phonetic distinctiveness when a speaker's utterance was masked by the feedback sound.

The present study examined whether the CAF could affect not only VOT productions of voiced and voiceless stop consonants (/d/ and /t/) but also perceptual identification of voiced (/d/)–voiceless (/t/) continuum stimuli varying in VOT. In addition, we focused on the relationship between the changes in phonetic production and perception, which may be caused by the CAF task, although their correlation was not found in a previous study using the TAF task (Lametti et al. 2014). We consider that the CAF task affects the motor-based speech processing but not the auditory-based speech processing, because speakers can notice that a

feedback phoneme is completely different from a produced phoneme and the auditory characterizations of a produced phoneme (/d/ or /t/) are not updated by the CAF task, while Patri et al. (2018) considered that the TAF modifies not only the motor-based but also the auditory-based speech processing. In addition, based on the suggestion of Mitsuya et al. (2014), voicing production changes caused by the CAF task would not be caused by an update of the auditory–motor internal models. Therefore, it is predictable that the effects of the CAF on phonetic production and perception are different from those of the TAF.

Regarding experimental design, Lametti et al. (2014) pointed out that it is important to take into consideration the effect of the selective auditory adaptation on phonetic perception, caused by listening repetitively to a particular phoneme sound during the TAF task, and distinguish its effect from the effect of motor functional changes on phonetic perception. In several studies using voicing perception, it was shown that the selective auditory adaptation caused by repetitive listening to voiced or voiceless stop consonants made the listeners more likely to perceive the voiced–voiceless continuum stimuli as voiceless or voiced stops (Eimas and Corbit 1973; Miller et al. 1983; Samuel 1982). In the present study, we had participants listen to (utter) both of voiced /d/ (voiceless /t/) and voiceless /t/ (voiced /d/) stop consonants in the CAF task, so that selective auditory adaptation would not occur. However, we could not rule out the possibility that listening to both voiced and voiceless stop consonants in the CAF task could affect voicing perception. Therefore, we also examined the effect of passive listening (PL) to the stimuli used in the CAF task on voicing perception in the experiment.

Methods

Figure 1 shows the flow of this experiment and a graphical representation of our cross-categorical auditory feedback (CAF) task procedure using voiced and voiceless syllable sounds, /da/ and /ta/. The experiment had two conditions: the auditory feedback (AF) and auditory adaptation (AA) conditions. The AF condition began with a non-perturbed production (NP) task in which the participants uttered /da/ and /ta/ without perturbation of auditory feedback. Then, the participants performed the first session of a perceptual identification task of /da/-/ta/ continuum stimuli. The first perceptual task was followed by the CAF task. Lastly, the second session of the perceptual identification task was conducted. In the AA condition, two sessions of phonetic identification task were conducted before and after the passive listening (PL) task with the stimuli used in the CAF task. The procedure for each task is described in detail below.

Participants

Seventeen native Japanese speakers, six females and 11 males (mean age 21.6 years, ranging from 19 to 26), participated. All participants gave informed consent before participation in the experiment. This study received approval from the Research Ethics Board of the Faculty of Information Science and Electrical Engineering, Kyushu University and was carried out in accordance with the latest version of the Declaration of Helsinki. Normal hearing was confirmed by measuring pure tone audiometric thresholds at octave intervals of 125–8000 Hz using an audiometer (RION, AA-M1C). All participants' thresholds were lower than 15 dB

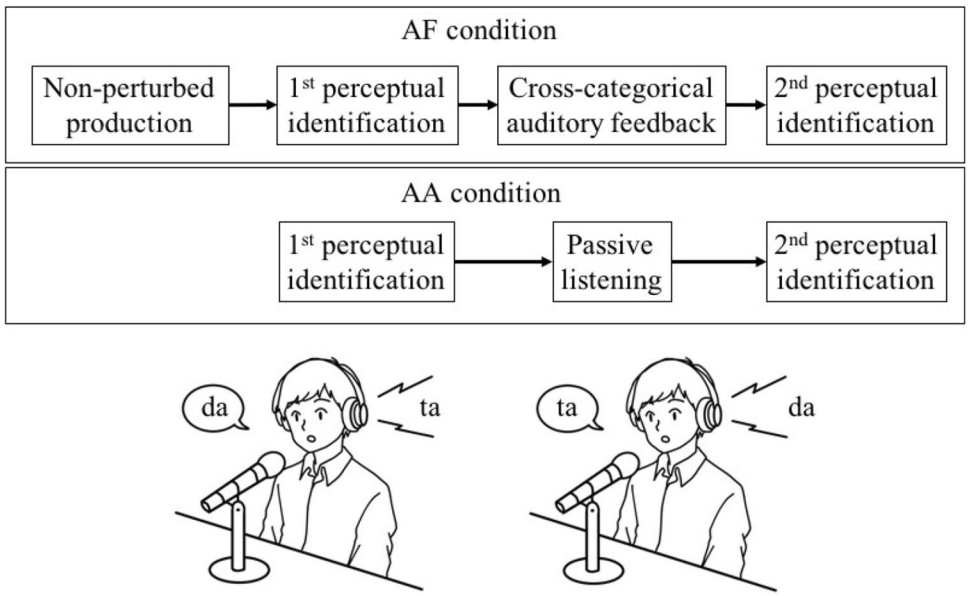

Fig. 1 Experimental flow of the present study and a graphic description of the cross-categorical auditory feedback task

at any frequency. All the participants performed the AF and AA conditions, with counterbalanced orders across participants. The AF and AA conditions were conducted on different days more than a week apart.

Stimuli

The stimuli for the CAF and the PL tasks were syllable sounds of /da/ and /ta/ recorded from each participant. Participants uttered /da/ and /ta/ syllables 30 times each before the first day of the experiment. For each syllable, as a feedback sound, we selected an utterance with VOT being closest to the mean of positive VOT (a voiced stop consonant /d/ had negative or positive VOT when the vibration of the vocal cords preceded or followed the consonant release and a voiceless stop consonant /t/ had a longer positive VOT than /d/). The speech stimuli were presented to both ears at a sound pressure level of approximately 85 dB in both the CAF and PL tasks, so that participants could not listen to their utterances in the CAF task. A broadband noise was presented at a sound pressure level of 50 dB throughout the CAF task to prevent participants from hearing their own utterances via bone conduction. This noise was also presented in the PL task to use the same auditory stimuli in the CAF and PL tasks.

The stimuli of the perceptual identification task were /da/-/ta/ continuum stimuli varying in VOT. These stimuli were composed of noise (consonant) and periodic (vowel) portions, which were created by a source-filter speech synthesizer (Klatt 1980) implemented in Praat 5.4.08 (Boersma and Weenink 2001). The source of the noise portion was turbulence noise, and the periodic portion had a voiced source with a fundamental frequency of 100 Hz. We made /da/-/ta/ continuum stimuli by varying the duration of the noise portion from 3 to 27 ms in 3-ms steps. The duration of the periodic portion was 140 ms. The first, fourth, and fifth formants started at stimulus onset and the frequencies were 800, 3300, and 3750 Hz without the transitions, respectively. The second and third formants started at the periodic portion onset and had a 40-ms frequency transition from 1600 to 1200 Hz and from 3000 to 2500 Hz, respectively. The stimuli were presented to each participant's right ear at a sound pressure level of 85 dB. The reason for presenting the auditory stimuli to the right ear only is that the brain networks in the left hemisphere were expected to be closely related to the effect of motor learning on speech perception and the perception of speech sounds, which were presented to the right ear, reflects mainly processing in the left hemisphere because of strong projections from the right ear to the auditory cortex of the left hemisphere (Sininger and Bhatara 2012).

A sound generation system (Tucker-Davis Technologies, System3), headphones (STAX, SR-407), and a headphone amplifier (STAX, SRM-006tS) were used to present the stimuli in the CAF, PL, and perceptual identification tasks. The sound pressure levels were measured using a Brüel and Kjær sound level meter (type 2260), a 1/2 inch condenser microphone (type 4192) and an artificial ear (type 4153).

Procedures

All the tasks in the AF and AA conditions were conducted in a sound-attenuated room. In the NP and CAF tasks, which were conducted in the AF condition, the participants uttered /da/ and /ta/ syllables 100 times each in random order, according to the visual instruction on a liquid crystal display (EIZO, FlexScan S2000). In the CAF task, the syllable sound, /ta/ or /da/, was presented almost simultaneously to the beginning of the participant's utterance of /da/ or /ta/ (Fig. 1). The delay between the participant's utterance and the presentation of speech stimuli was less than 10 ms. A microphone (AKG, D7S), a sound amplifier (EDIROL, UA-3D) and PCMCIA audio interface (Echo Audio, Indigo IO) attached to a personal computer were used to record the utterance in the NP and CAF tasks and detect the utterance onset in the CAF task. The utterance was digitized at 48 kHz and the digitized signals were directly delivered to MATLAB (Mathworks, v.7.0.1) using Data Acquisition Toolbox (Mathworks, v.2.5.1). In the PL task, the participants listened to the syllable sound, /da/ or /ta/, while they were visually presented with the syllable, /ta/ or /da/. Each syllable was presented 100 times each in random order.

For the perceptual identification task of /da/-/ta/ continuum stimuli, we used a one-interval, two-alternative forced-choice task. The participants were presented with one of the /da/-/ta/ continuum stimuli randomly on each trial and were asked to judge whether the presented stimulus was /da/ or /ta/. Each stimulus was presented 10 times and the total number of trials was 90 in the main session. Before the first session of the task, each participant performed a practice session in which the endpoint stimuli of the continuum stimuli were randomly presented. Before the practice session, participants were instructed that the stimulus of 3-ms VOT was /da/ and that of 27 ms VOT was /ta/. The two stimuli were presented 10 times each for a total of 20 trials in the practice session. The practice session continued until correct responses to the endpoint stimuli reached 90%.

Results

Production

To examine the changes in voicing production, we compared VOTs of /da/ and /ta/ production during the NP and CAF tasks. We determined the VOT by measuring the time interval between onsets of the noise (consonant) and periodic

(vowel) portions of an utterance. Regarding the /da/ utterances, the overall mean proportion of positive VOT was approximately 0.8 during the NP task and 0.7 during the CAF task, with no significant difference between the tasks. As the number of utterances with negative VOT was very small in several participants, we did not analyze further them. On the other hand, all the /ta/ utterances had positive VOT. The individual mean positive VOTs of /da/ and /ta/ utterances were calculated for each task. Figure 2a, b show the overall mean VOT of /da/ and /ta/, respectively, during the NP and CAF tasks. We conducted paired *t* tests separately for /da/ and /ta/ and found that the VOT of /ta/ production during the CAF task was longer than that during the NP task [$t(16) = 3.49$, $p < 0.01$, $d = 0.90$], while there was no significant difference between the VOT of /da/ production in the two tasks [$t(16) = 0.90$, $p = 0.38$, $d = 0.77$]. In addition, the individual VOT difference between /da/ and /ta/ production was calculated. Figure 2c shows the overall means in the NP and CAF tasks. A paired *t* test revealed that the CAF task caused the participants to increase the difference between /da/ and /ta/ production [$t(16) = 3.45$, $p < 0.01$, $d = 0.89$].

Perception

For data analysis of the /da/-/ta/ perceptual identification task, the individual proportions of /ta/ responses for the nine continuum stimuli were calculated before and after the CAF and PL tasks. In addition, the individual VOT boundaries and slopes of the /ta/ response functions before and after each task were estimated using the following logistic function:

$$P(\mathrm{VOT}) = 1/\left\{1 + \mathrm{e}^{-\beta(\mathrm{VOT}-\alpha)}\right\}. \quad (1)$$

P(VOT) indicates the proportions of /ta/ responses at a given VOT value. The parameters α and β indicate the VOT value at $P(\mathrm{VOT}) = 0.5$, i.e., the VOT boundary and the slope of the logistic function, which correspond to the precision in voiced–voiceless perception, respectively. This function was fitted to the individual proportion data by the iteratively reweighted least squares method. For the statistical analysis of perception data, we removed one participant's data because the slope value was more than 3SD from the overall mean in the perceptual identification task after the CAF task.

Figure 3a shows mean identification functions of the /ta/ response before and after the CAF and PL tasks, which were created based on overall mean values of slope and the VOT boundary in each task. Figure 3b shows the overall mean slopes before and after the CAF and PL tasks. We conducted a two-way repeated-measures analysis of variance (ANOVA) on the slopes with session (before vs. after) and task (CAF vs. PL) as within-participant factors. There was no significant main effect of session [$F(1,15) = 0.11$, $p = 0.74$, $\eta^2 = 0.01$] or task [$F(1,15) = 0.26$, $p = 0.61$, $\eta^2 = 0.02$]. The interaction between session and task was significant [$F(1,15) = 5.14$, $p = 0.04$, $\eta^2 = 0.26$]. Simple main effect analyses revealed that the slope was significantly steepened by the CAF task [$F(1,15) = 5.30$, $p = 0.04$, $\eta^2 = 0.26$], while the slope was not changed by the PL task [$F(1,15) = 1.90$, $p = 0.19$, $\eta^2 = 0.11$]. Figure 3c shows the overall mean VOT boundaries before and after the CAF and PL tasks. A two-way repeated-measures ANOVA on the boundaries found no significant main effect of session [$F(1,15) = 2.83$, $p = 0.11$, $\eta^2 = 0.16$] or task [$F(1,15) = 0.01$, $p = 0.97$, $\eta^2 = 0.00$]. The interaction between session and task was not significant either [$F(1,15) = 3.06$, $p = 0.14$, $\eta^2 = 0.14$].

Correlation between production and perception results

Finally, we investigated the correlation between the changes in voicing production and perception. We calculated the increases in the VOT difference between /da/ and /ta/

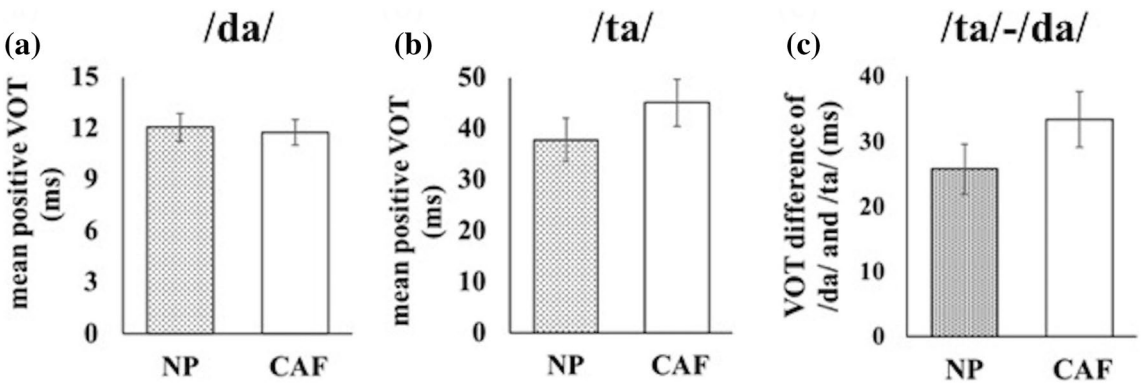

Fig. 2 Comparisons of /d/ and /t/ utterance voice-onset times (VOTs) between the non-perturbed production (NP) and cross-categorical auditory feedback (CAF) tasks. Overall mean VOT of /d/ (**a**) and /t/ (**b**) during the NP and CAF tasks. **c** Overall mean of the VOT difference between /da/ and /ta/ production in the NP and CAF tasks. Error bars represent standard errors of mean values

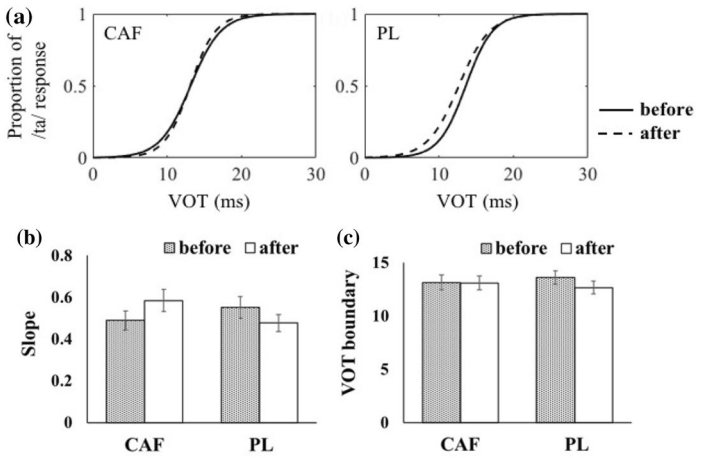

Fig. 3 Comparisons of /da/-/ta/ categorization results before and after cross-categorical auditory feedback (CAF) and passive listening (PL) tasks. **a** Overall mean identification functions of /t/ to the /da/-/ta/ continuum stimuli before and after the CAF and PL tasks, which were created based on overall mean values of slope and the voice-onset time (VOT) boundary in each task. Solid and dotted lines indicate the identification functions before and after the two tasks. **b** Overall mean slopes and **c** overall mean VOT boundaries before and after the CAF and PL tasks. Error bars represent standard errors of mean values

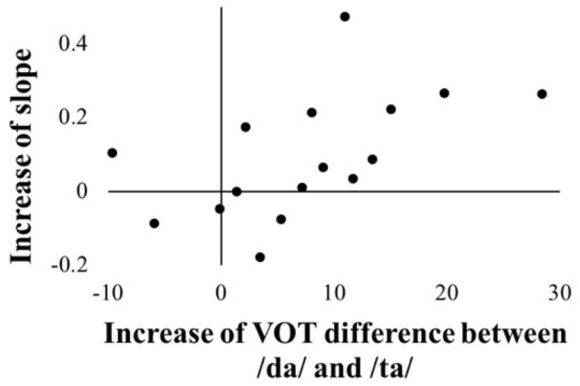

Fig. 4 Scatterplot of the increases in perceptual slope and voice-onset time (VOT) difference between /da/ and /ta/ production from before to after the cross-categorical auditory feedback (CAF) task

production from the NP task to the CAF task and the identification slope from before to after the CAF task for each participant. Figure 4 shows a scatterplot of the increases in VOT difference and slope. Spearman's correlation analysis showed a significant positive correlation [$r(14) = 0.64$, $p < 0.01$].

Discussion

The present study examined whether the CAF task altered not only voicing production but also perception. Concerning voicing production, we found that the CAF task caused the speakers to lengthen the VOT of /ta/ production and increased the VOT difference between /da/ and /ta/ production. This finding is consistent with the results of a previous study using the CAF task (Mitsuya et al. 2014), although we could not replicate the shortening of VOT of voiceless production.

Concerning voicing perception, we found that the identification slope was significantly steepened after the CAF task but not after the PL task. In addition, a significant correlation between increases in the VOT difference between /da/ and /ta/ productions and in the slope of the /da/-/ta/ identification function was found. These results are inconsistent with the results of the previous studies using the TAF task in two respects (Lametti et al. 2014; Shiller et al. 2009). Firstly, our CAF task affected the precision of phonetic identification, while the TAF shifted the phonetic boundary in previous studies. Secondly, we found a significant correlation between the changes in phonetic production and perception, although Lametti's study did not find their significant correlations.

A key question is what functional changes caused by the CAF task affected voicing perception. We addressed this question focusing on the difference between our CAF task and the TAF task used in the previous studies (Lametti et al. 2014; Shiller et al. 2009). Concerning the effect of the TAF on phonetic perception, Patri et al. (2018) suggested that the updates of the auditory–motor internal models and auditory characterizations of the produced phoneme, caused by the TAF, altered both the auditory-based and the motor-based speech processing for phonetic perception. On the other hand, we consider that the impact of the CAF on voicing perception was mainly due to the update of motor-based speech processing because auditory characterization of the produced phoneme cannot be updated by the CAF as described in the "Introduction" section. Such a difference

between functional changes caused by the CAF and TAF may be the main reason for the inconsistency of the correlation results of our study and Lametti's study.

It is conceivable that the voicing production changes observed in the present study resulted not from the update of the auditory–motor internal models, but rather from feedforward control to maintain phonetic distinctiveness as suggested by Mitsuya et al. (2014). In addition, we speculate that feedforward control is also closely related to voicing perception changes. Although we cannot fully explain how feedforward control causes an improvement in the precision of perceptual identification of voiced–voiceless continuum stimuli, the significant correlation between phonetic production and perception changes, found in the present study, corresponds to the previous findings that speakers who more acutely discriminate a phonetic contrast produce that contrast more distinctly (e.g., Perkell et al. 2004). Concerning the relationship between voicing production and perception, our present findings are consistent with the results of previous lesion studies in which the lesions in brain regions involved with speech production obscured the difference between voiced and voiceless production (Blumstein et al. 1980; Ivry and Gopal 1993) and reduced perceptual precision of the voiced–voiceless continuum varying in VOT (Ackermann et al. 1997; Basso et al. 1977).

In the following section, we discuss the present findings considering a current theory regarding speech processing in the brain (Rauschecker 2011; Rauschecker and Scott 2009). It is assumed that the processing in the dorsal auditory stream mediates the relationship between auditory and motor representations of speech during speech production and perception and that the ventral stream is involved in processing speech sounds for comprehension on the basis of auditory information. Based on this assumption, the dorsal and ventral streams may be closely related to the possible effects of the updates of auditory–motor internal models and a produced phoneme's auditory characterization, respectively, which were caused by the TAF, on phonetic production and perception (Patri et al. 2018). On the other hand, we propose a possibility that the brain networks mediating the relationships between phonemes and articulatory movements are also connected to phonetic perception based on our suggestion that a phonetic production change itself, caused by the CAF, alters phonetic perceptual processing.

We obtained some meaningful results to discuss the interactions between speech perception and production, but there are several limitations in our study. First, one might find it problematic that the auditory stimuli were presented to the right ear only in the /da/-/ta/ identification task while those were heard binaurally in the CAF and PL tasks. However, we believe that the results would be the same as the present results if the stimuli were presented to both ears in the identification task. This is because information from the right ear is preferentially used for speech perception in listening with both ears (Sininger and Bhatara 2012). Second, the effect of the CAF on the slope of /da/-/ta/ identification function was statistically significant but so small. A possible reason for this result is that the individual difference of voicing perception change induced by the CAF was very large. In fact, the slope became less steep after than before the CAF task in several participants. Therefore, the effect of the CAF on the slope might have been not very clear on the whole. However, what is important in the present study's result is that voicing perception change correlated with its production change.

In addition to the hypothesis for mechanisms underlying phonetic perception focusing on the connection between speech production and perception systems (Liberman et al. 1967; Liberman and Mattingly 1985; Stevens and Halle 1967), there is another hypothesis that nonlinear characteristics in auditory processing of speech sounds provide critical information for phonetic perception (Diehl 2008; Holt et al. 2004; Stevens 1989). Several studies suggested that nonlinear auditory temporal processing provides useful information for perceptual identification of voiced–voiceless contrasts in stop consonants (Pisoni 1977; Steinschneider et al. 2004; Tamura et al. 2018). Based on the findings of the present study and previous studies examining the auditory mechanism for voicing perception, it is conceivable that both auditory-based speech processing and motor-based speech processing contribute to phonetic perception as described in several recent studies (Devlin and Aydelott 2009; Hickok et al. 2011; Patri et al. 2018; Schwartz et al. 2012). Therefore, in future studies, it will be necessary to clarify the difference between the roles of auditory-based and motor-based speech processing to fully understand the overall mechanism underlying phonetic perception.

Acknowledgements This research was supported by JSPS KAKENHI Grant numbers JP18J10654, JP25240023.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Ackermann H, Gräber S, Hertrich I, Daum I (1997) Categorical speech perception in cerebellar disorders. Brain Lang 60:323–331

Basso A, Casati G, Vignolo LA (1977) Phonemic identification defect in aphasia. Cortex 13:85–95

Blumstein SE, Cooper WE, Goodglass H, Statlender S, Gottlieb J (1980) Production deficits in aphasia: a voice-onset time analysis. Brain Lang 9:153–170

Boersma P, Weenink D (2001) Praat, a system for doing phonetics by computer. Glot Int 5:341–345

Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M (2013) Automatic phoneme category selectivity in the dorsal auditory stream. J Neurosci 33:5208–5215

Devlin JT, Aydelott J (2009) Speech perception: motoric contributions versus the motor theory. Curr Biol 19:198–200

Diehl RL (2008) Acoustic and auditory phonetics: the adaptive design of speech sound systems. Philos Trans R Soc B 363:965–978

Eimas PD, Corbit JD (1973) Selective adaptation of linguistic feature detectors. Cogn Psychol 4:99–109

Evans S, Davis MH (2015) Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. Cereb Cortex 25:4772–4788

Hickok G, Costanzo M, Capasso R, Miceli G (2011) The role of Broca's area in speech perception: evidence from aphasia revisited. Brain Lang 119:214–220

Holt LL, Lotto AJ, Diehl RL (2004) Auditory discontinuities interact with categorization: implications for speech perception. J Acoust Soc Am 116:1763–1773

Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. Science 279:1213–1216

Houde JF, Jordan MI (2002) Sensorimotor adaptation of speech I: compensation and adaptation. J Speech Lang Hear R 45:295–310

Ivry RB, Gopal HS (1993) Speech production and perception in patients with cerebellar lesions. In: Meyer DE, Kornblum S (eds) Attention and performance XIV. MIT Press (A Bradford Book), Cambridge, pp 771–802

Jones JA, Munhall KG (2000) Perceptual calibration of F0 production: evidence from feedback perturbation. J Acoust Soc Am 108:1246–1251

Jones JA, Munhall KG (2005) Remapping auditory-motor representations in voice production. Curr Biol 15:1768–1772

Klatt DH (1980) Software for a cascade/parallel formant synthesizer. J Acoust Soc Am 67:971–995

Lametti DR, Rochet-Capellan A, Neufeld E, Shiller DM, Ostry DJ (2014) Plasticity in the human speech motor system drives changes in speech perception. J Neurosci 34:10339–10346

Lee YS, Turkeltaub P, Granger R, Raizada RD (2012) Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. J Neurosci 32:3942–3948

Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cognition 21:1–36

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. Psychol Rev 74:431–461

Lisker L, Abramson AS (1964) A cross-language study of voiced-voiceless in initial stops: acoustical measurements. Word 20:384–422

Miller JL, Connine CM, Schermer TM, Kluender KR (1983) A possible auditory basis for internal structure of phonetic categories. J Acoust Soc Am 73:2124–2133

Mitsuya T, MacDonald EN, Munhall KG (2014) Temporal control and compensation for perturbed voiced-voiceless feedback. J Acoust Soc Am 135:2986–2994

Ostry DJ, Gribble PL (2016) Sensory plasticity in human motor learning. Trends Neurosci 39:114–123

Patri JF, Perrier P, Schwartz JL, Diard J (2018) What drives the perceptual change resulting from speech motor adaptation? Evaluation of hypotheses in a Bayesian modeling framework. PLoS Comput Biol 14:1–38

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, Zandipour M (2004) The distinctness of speaker's production of vowel contrasts is related to their discrimination of the contrasts. J Acoust Soc Am 116:2338–2344

Pisoni DB (1977) Identification and discrimination of the relative onset time of two component tones: implications for voiced-voiceless perception in stops. J Acoust Soc Am 61:1352–1361

Pulvermüller F, Huss M, Kherif F, del Prado Martin FM, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci USA 103:7865–7870

Rauschecker JP (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res 271:16–25

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724

Samuel AG (1982) Phonetic prototypes. Percept Psychophys 31:307–314

Schomers MR, Pulvermüller F (2016) Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. Front Hum Neurosci 10:1–18

Schwartz JL, Basirat A, Ménard L, Sato M (2012) The perception-for-action-control theory (PACT): a perceptuo-motor theory of speech perception. J Neurolinguist 25:336–354

Shiller DM, Sato M, Gracco VL, Baum SR (2009) Perceptual recalibration of speech sounds following speech motor learning. J Acoust Soc Am 125:1103–1113

Sininger YS, Bhatara A (2012) Laterality of basic auditory perception. Laterality 17:129–149

Skipper JI, Devlin JT, Lametti DR (2017) The hearing ear is always found close to the speaking tongue: review of the role of the motor system in speech perception. Brain Lang 164:77–105

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA III (2004) Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. Cereb Cortex 15:170–186

Stevens KN (1989) On the quantal nature of speech. J Phon 17:3–45

Stevens KN, Halle M (1967) Remarks on analysis by synthesis and distinctive features. In: Walthen-Dunn W (ed) Models for the perception of speech and visual form. The MIT Press, Massachusetts, pp 88–102

Tamura S, Ito K, Hirose N, Mori S (2018) Psychophysical boundary for categorization of voiced–voiceless stop consonants in native Japanese speakers. J Speech Lang Hear R 61:789–796

Villacorta VM, Perkell JS, Guenther FH (2007) Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. J Acoust Soc Am 122:2306–2319

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702