



Repetitive Exposure to Orofacial Somatosensory Inputs in Speech Perceptual Training Modulates Vowel Categorization in Speech Perception

Takayuki Ito^{1,2*} and Rintaro Ogane^{1,2}

¹ Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France, ² Haskins Laboratories, New Haven, CT, United States

OPEN ACCESS

Edited by:

Christopher Carignan,
University College London,
United Kingdom

Reviewed by:

Donald Derrick,
University of Canterbury, New Zealand
Lucie Menard,
Université du Québec à Montréal,
Canada

*Correspondence:

Takayuki Ito
takayuki.ito@gipsa-lab.grenoble-inp.fr

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 19 December 2021

Accepted: 24 March 2022

Published: 26 April 2022

Citation:

Ito T and Ogane R (2022)
Repetitive Exposure to Orofacial
Somatosensory Inputs in Speech
Perceptual Training Modulates Vowel
Categorization in Speech Perception.
Front. Psychol. 13:839087.
doi: 10.3389/fpsyg.2022.839087

Orofacial somatosensory inputs may play a role in the link between speech perception and production. Given the fact that speech motor learning, which involves paired auditory and somatosensory inputs, results in changes to speech perceptual representations, somatosensory inputs may also be involved in learning or adaptive processes of speech perception. Here we show that repetitive pairing of somatosensory inputs and sounds, such as occurs during speech production and motor learning, can also induce a change of speech perception. We examined whether the category boundary between /ε/ and /a/ was changed as a result of perceptual training with orofacial somatosensory inputs. The experiment consisted of three phases: Baseline, Training, and Aftereffect. In all phases, a vowel identification test was used to identify the perceptual boundary between /ε/ and /a/. In the Baseline and the Aftereffect phase, an adaptive method based on the maximum-likelihood procedure was applied to detect the category boundary using a small number of trials. In the Training phase, we used the method of constant stimuli in order to expose participants to stimulus variants which covered the range between /ε/ and /a/ evenly. In this phase, to mimic the sensory input that accompanies speech production and learning in an experimental group, somatosensory stimulation was applied in the upward direction when the stimulus sound was presented. A control group (CTL) followed the same training procedure in the absence of somatosensory stimulation. When we compared category boundaries prior to and following paired auditory-somatosensory training, the boundary for participants in the experimental group reliably changed in the direction of /ε/, indicating that the participants perceived /a/ more than /ε/ as a consequence of training. In contrast, the CTL did not show any change. Although a limited number of participants were tested, the perceptual shift was reduced and almost eliminated 1 week later. Our data suggest that repetitive exposure of somatosensory inputs in a task that simulates the sensory pairing which occurs during speech production, changes perceptual system and supports the idea that somatosensory inputs play a role in speech perceptual adaptation, probably contributing to the formation of sound representations for speech perception.

Keywords: somatosensory stimulation, perceptual adaptation, multisensory integration, production-perception link, auditory representation

INTRODUCTION

Speech perception is auditory in nature but it is also an interactive process involving other sensory inputs. For example, visual information coming from a speaker's face helps in the identification of speech sounds in a noisy environment (Sumbly and Pollack, 1954). Incongruent visual information from facial movements likewise affects speech perception (McGurk and MacDonald, 1976). Recent studies have demonstrated that somatosensory inputs also contribute to the perception of speech. When air-puffs, similar to those associated with a plosive speech sound (such as /p/), were presented to the skin, perception was biased in the direction of the corresponding sound (Gick and Derrick, 2009). When somatosensory stimulation using facial skin deformation was applied in conjunction with the speech sounds, vowel perception was systematically biased (Ito et al., 2009). In a vowel identification task on a "head/had" continuum, the presented vowels were perceived more as "head" when an upward skin stretch was applied, more as "had" when the skin stretch was downward, and there was no effect with backward skin stretch. A similar effect has been observed in both children and adults using a vowel continuum between /e/ and /ø/ (Trudeau-Fisette et al., 2019). When the skin stretch was backward, the presented sounds were perceived more as /e/, a vowel in which lip spreading is involved in production. A somatosensory influence on perception is not limited to vowel categorization, but is also observed in word segmentation in lexical processing (Ogane et al., 2020). The segmentation boundary changed depending on the placement of somatosensory stimulation in relation to the key vowel in a test phrase. While these studies suggest a potential role of somatosensory inputs in speech perception, the specific contribution of the somatosensory system is unknown.

Given that orofacial somatosensory inputs normally provide articulatory information in the context of speech production (Johansson et al., 1988; Ito and Gomi, 2007; Ito and Ostry, 2010), somatosensory effects on speech perception may be production related. This idea was initially proposed in the Motor Theory of Speech Perception (Lieberman et al., 1967), and extended in the Direct Realist perspective (Fowler, 1986) and the Perception-for-Action-Control theory (Schwartz et al., 2012). The possible contribution of the sensorimotor system to perception has mostly focused on the motor system. For example, activity in brain motor areas has been observed during speech perception (Wilson et al., 2004; Skipper et al., 2005), and the perception of speech sounds can be modulated by applying transcranial magnetic stimulation to the premotor cortex (Meister et al., 2007; D'Ausilio et al., 2009). At a behavioral level, when speech articulation is simultaneous with listening, the perception of speech sounds is altered (Sams et al., 2005; Mochida et al., 2013; Sato et al., 2013). However, speech motor outflow always occurs in conjunction with correlated somatic input. While somatosensory function might be considered part of motor system, the somatosensory system may work independently in the perception of speech sounds since there is a direct influence and interaction between the somatosensory and auditory system in situations other than speech perception (Foxe et al., 2000; Beauchamp et al.,

2008). Thus, investigating somatosensory function in speech perception may be important in clarifying the link between speech production and perception.

The contribution of somatosensory inputs to speech perception has been examined in the context of motor learning. Previous studies showed that adapting to different external environments during production changes the vowel category boundary (Nasir and Ostry, 2009). Similar perceptual changes have been reported in studies of adaptation to altered auditory feedback (Shiller et al., 2009; Lametti et al., 2014). Although both motor outputs and somatosensory inputs are involved in the speech motor learning tasks used in these previous studies, Ohashi and Ito (2019) specifically demonstrated that somatosensory inputs on their own can contribute to the recalibration of perception. That study applied additional somatosensory stimulation during adaptation to altered auditory feedback and assessed changes to the category boundary of fricative consonants. They observed perceptual recalibration in conjunction with somatosensory stimulation, suggesting that repetitive exposure to somatosensory inputs during learning can be a key to changing or recalibration of the speech perceptual representation.

In addition to motor learning, repetitive exposure to sensory stimuli also induces changes to sensory processing. In the case of speech, the phonetic boundary between two neighboring speech sounds can be biased away from the one that is repetitively presented as an adapter in training, which is known as selective adaptation (Eimas and Corbit, 1973). Similar effects can be seen in visual speech perception (Jones et al., 2010). This type of sensory adaptation has been frequently investigated in non-linguistic processing. In the visual domain, after looking at a high-contrast visual image, a low-contrast portion of a test image briefly appears invisible (e.g., Kohn, 2007). Similarly, after prolonged observation of a waterfall, an illusory upward motion can be induced when observing a static image (Mather et al., 1998). This effect has also been demonstrated in multisensory environments, including selective adaptation in audio-visual speech (Roberts and Summerfield, 1981; Saldaña and Rosenblum, 1994; Dias et al., 2016). In case of ambiguous speech sounds, visual information from speech movements changes auditory perception (Bertelson et al., 2003). Speech sounds also change visual speech perception (Baart and Vroomen, 2010). If somatosensory inputs contribute to the formulation or calibration of speech perceptual representations, repetitive exposure to orofacial somatosensory stimulation, such as occurs normally in conjunction with speech production and learning, may recalibrate the representation of speech sounds. If this adaptive change persists following perceptual training, then training with somatosensory stimulation may potentially be used as a tool for speech training and rehabilitation.

The present study examined whether repetitive exposure to orofacial somatosensory stimulation during a speech perception task changes the perception of speech sounds. To test this idea, we here focused on the category boundary between the vowels /ε/ and /a/ and applied orofacial somatosensory stimulation, specifically facial skin deformation, as used in previous studies (Ito et al., 2009; Trudeau-Fisette et al., 2019;

Ogane et al., 2020). The use of orofacial somatosensory stimulation is premised on the assumption that skin receptors provide kinesthetic information (Johansson et al., 1988; Ito and Gomi, 2007; Ito and Ostry, 2010). Given that somatosensory stimulation involving facial skin deformation changed the category boundary between “head” and “had” in on-line manner (Ito et al., 2009), training with the same auditory-somatosensory pairing may change or recalibrate the vowel category boundary in purely auditory perceptual tests. We carried out perceptual training paired with somatosensory stimulation and assessed changes to the category boundary. It might be expected, based on our prior work using a simple perceptual classification task (Ito et al., 2009), that upward skin stretch during vowel identification on a / ϵ / to / a / continuum would bias perception toward / ϵ /. However, if the effect of the training task on perception is similar to selective adaptation mentioned above, training might be accompanied by a perceptual shift toward / a /. Either perceptual change would suggest that the somatosensory system contributes to the link between speech production and perception, and that somatosensory inputs can help in the processing of speech sounds in ambiguous situations.

MATERIALS AND METHODS

Participants

Thirty native speakers of French participated in the experiment. The participants were all healthy young adults who reported normal hearing. All participants signed informed consent forms approved by the Local Ethical Committee of the University Grenoble Alpes (CERNI: Comité d’Ethique pour les Recherches non Interventionnelles: Avis-2015-03-03-62 or CERGA: Comité d’Ethique pour la Recherche, Grenoble Alpes: Avis-2018-12-11-4).

Auditory Stimulation

We focused on vowel categorization using an / ϵ / to / a / continuum, based on a previous study which showed a clear somatosensory effect on speech perception (Ito et al., 2009). These vowels were followed by the / f / sound which is associated with a closing movement after the vowel production. The stimulus continuum was synthesized by using an iterative Burg algorithm for estimating spectral parameters. The procedure involved shifting the first (F1) and the second (F2) formant frequencies in equal steps from values observed for / ϵ / to those associated with / a /. The stimulus sound was recorded by male speaker of French. The first and second formant values for the endpoint stimuli were 561 Hz and 1630 Hz for / ϵ /, and 712 Hz and 1203 Hz for / a /. A forty-six-step continuum was produced for the adaptive testing procedure used in Baseline and Aftereffect tests; a subset of these stimuli was selected for use in perceptual training (see below).

Somatosensory Stimulation

We used facial skin stretch applied by a robotic device to produce somatosensory stimulation (Phantom 1.0; SensAble Technologies). The experimental setup is shown in **Figure 1A**.

Plastic tabs (2 cm \times 3 cm) were attached to the skin lateral to the oral angle on each side of the face. These tabs were connected to the robotic device through thin wires. The wires were supported by wire supports to avoid contact with the facial skin. The skin was stretched when the robotic device applied force to the wires. The temporal profile of the applied force was a single cycle of a 3-Hz sinusoid with 2N peak force (see **Figure 1B**). Based on the previous finding that the upward skin stretch induced a relatively large change in vowel categorization judgments between “head” and “had” (Ito et al., 2009), we applied the skin stretch in an upward direction.

Perceptual Test and Adaptation Training

The main test was consisted of three phases: Baseline, Training, and Aftereffect (see **Figure 1C**). In all three phases, an identification test using the vowels / ϵ / and / a / was involved. The stimuli were presented through head-phones at a comfortable volume. On each trial, participants were asked to identify whether the sound was / ϵf / or / $a f$ / by pressing a key on a keyboard.

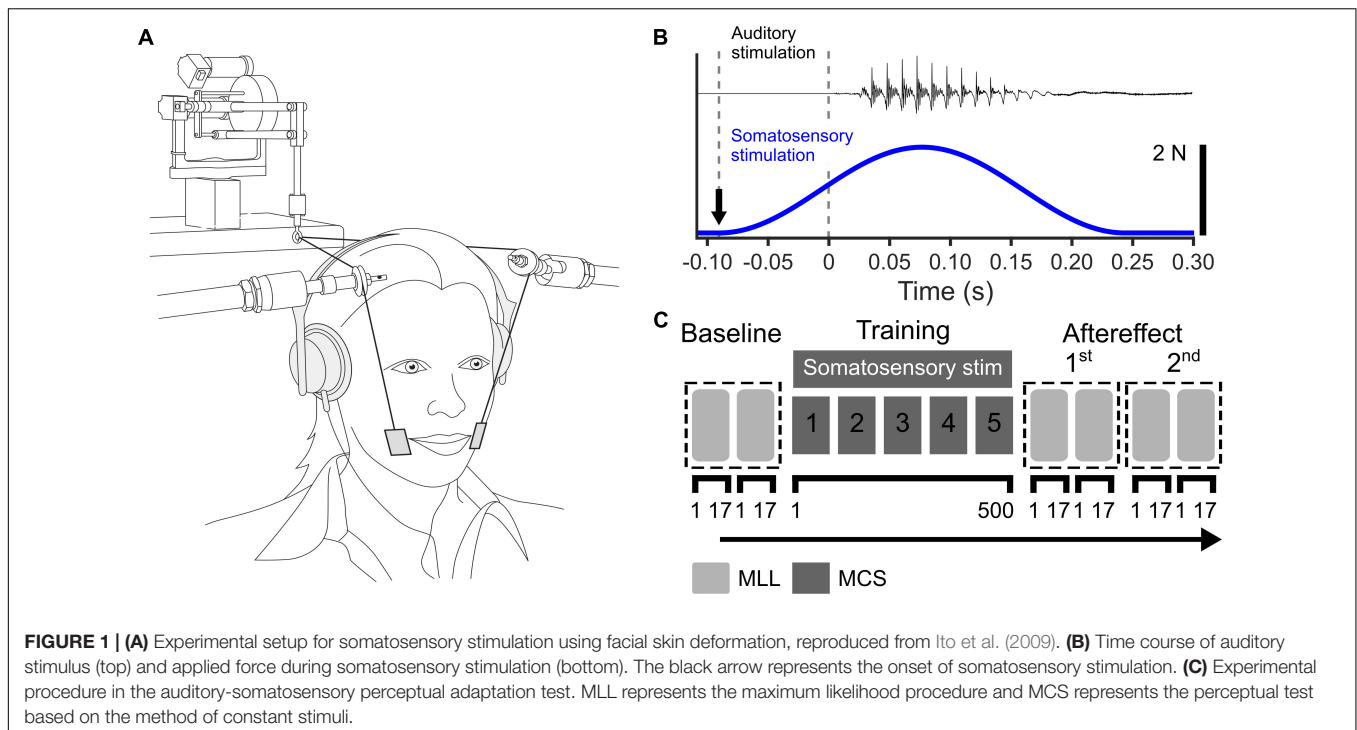
In the main perceptual training portion of the study, the method of constant stimuli (MCS) was used in order to expose participants to values between / ϵ / and / a / evenly during the training. We used 10 of 46 steps on the / ϵ / to / a / continuum (Nos. 1, 6, 11, 16, 21, 26, 31, 36, 41, and 46) and presented them 10 times each in pseudo-random order. Each training block consisted of 100 trials. This was repeated 5 times. In total, 500 stimuli were presented. For the experimental group which received somatosensory training (SOMA), somatosensory stimulation was applied on each trial. The temporal relationship between the sound stimulus and somatosensory stimulation is shown in **Figure 1B**. For the control group (CTL), we carried out the same training including the setup of the robot, but in absence of somatosensory stimulation.

In Baseline and Aftereffect tests, we used an adaptive method based on the maximum-likelihood (MLL) procedure to estimate the vowel category boundary (Shen and Richards, 2012). The benefit of this procedure is its ability to estimate the psychometric function and the associated category boundary with a relatively small number of responses in comparison to other conventional methods such as MCS. However, sounds near to the perceptual boundary are primarily tested. In this procedure, the auditory test stimulus on each trial is determined in an adaptive fashion based on the stimulus that provides the most information about the shape of the psychometric function. All stimuli on the forty-six-step continuum were used in this procedure. Each of the perceptual tests consisted of four 17-trial blocks. The first two blocks of the Baseline phase were removed from the analysis as familiarization trials for the identification task.

In order to examine if the effect of paired auditory-somatosensory training persisted 1 week later, we also repeated the Post-test using the same procedure as in the Aftereffect phase, based on MLL procedure. Five of 15 participants participated in the Post-test.

Data Analysis

We calculated the probability that the participant identified the presented vowel as / a /. We estimated the psychometric



function for each 17-trial block of the MLL procedure (Baseline, Aftereffect and Post-test) and for each 100-trial block of the MCS procedure (Training), and obtained estimates of the category boundary as the 50% value of the psychometric function. The baseline value for the category boundary was obtained by averaging the two blocks of the Baseline phase. In the Aftereffect and Post-test phases, we also averaged separately the first two (1st set) and the second two blocks (2nd set). The obtained category boundaries were normalized by dividing by the baseline boundary value.

To examine whether the category boundary changed following perceptual training, we applied a one sample *t*-test to the normalized perceptual boundary immediately following training (average of the first two blocks of the Aftereffect). This normalized perceptual boundary was also compared between control and somatosensory training groups using a Linear Mixed-Effects (LME) Models analysis with nlme package in R (Pinheiro et al., 2022). In the LME model including the following analyses, participants were always considered as a random effect.

We also applied a LME analysis to evaluate whether the perceptual boundary changed over the course of training (Training phase). Fixed factors were groups (CTL and SOMA) and blocks (1, 2, 3, 4, and 5). A separate one-sample *t*-test was also used to examine whether the category boundary averaged over the course of training was different than baseline.

The LME analysis was likewise used to evaluate the possible presence of persistent effect at a one-week delay. For this evaluation, we first compared changes between the 1st and 2nd sets in the Aftereffect phase, with groups (CTL and SOMA) and sets (1st and 2nd) as fixed factors. *Post hoc*

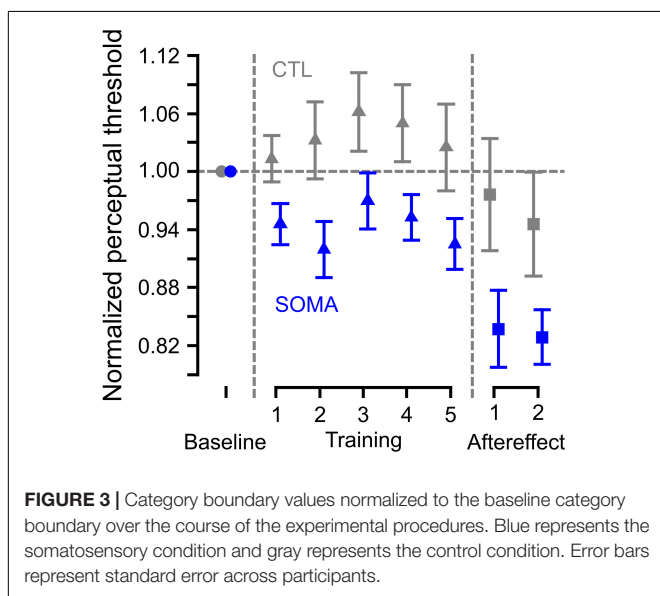
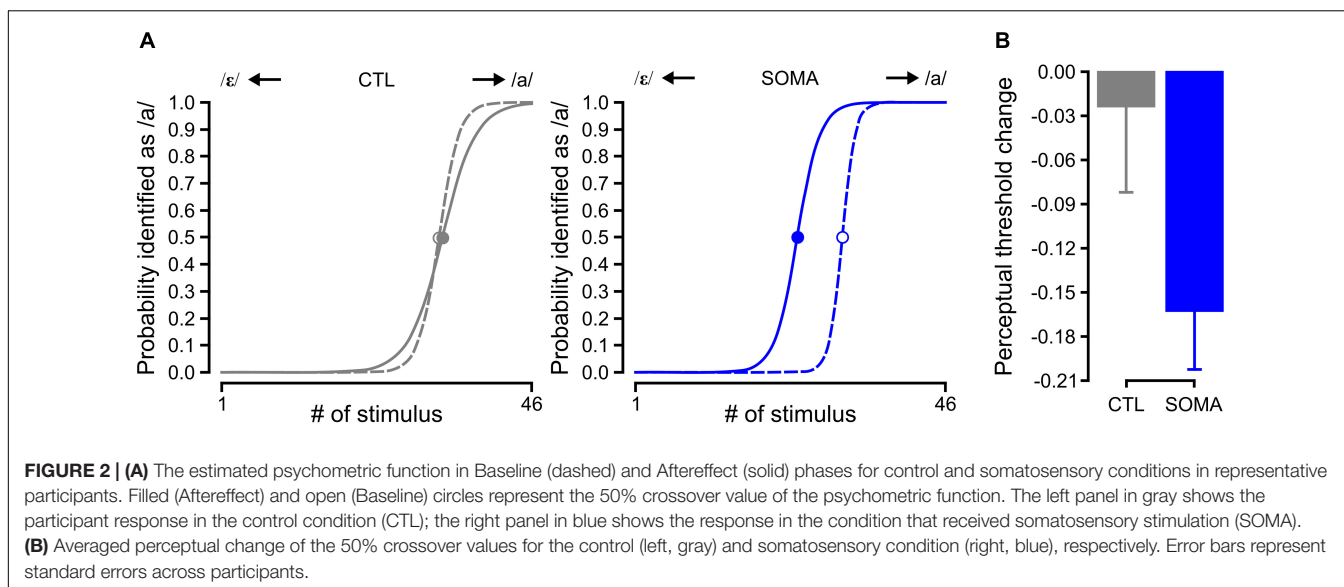
tests with Bonferroni correction were carried out to compare all possible combinations using the multcomp package in R (Hothorn et al., 2008). Second, we compared the Aftereffect and Post-test phases in the five participants that completed both. In this analysis, we extracted the category boundaries for these participants in the Baseline, Aftereffect and Post-test measures and calculated separately the normalized boundary in the Aftereffect and Post-test, as described above. We used a LME analysis to assess whether the normalized boundary was different in the Aftereffect and Post-test measures. Fixed factors in this analysis were phases (Aftereffect and Post-test) and sets (1st and 2nd).

RESULTS

Shift of Category Boundary Due to the Training

Figure 2A shows representative results for the estimated psychometric function prior to and following training in the two conditions (CTL and SOMA). As shown here, the category boundary shifted in the direction of /ε/ following training with somatosensory stimulation (SOMA, solid blue line in the right panel of Figure 2A), indicating that the participants perceived /a/ more than /ε/ as an aftereffect. This shift was not observed following training in the control condition (CTL, solid gray line in the left panel of Figure 2A). Averaged perceptual changes with standard errors are shown in Figure 2B.

The amplitude of the shift was significantly different from zero [-0.163 ± 0.040 , average \pm s.e., $t(14) = -4.08$, $p < 0.005$] after the training with somatosensory stimulation



(SOMA). In the control condition (CTL), the magnitude of the shift was not different than zero [-0.023 ± 0.058 , average \pm s.e., $t(14) = -0.41$, $p > 0.6$]. A comparison between groups using a LME analysis also showed a significant effect [$\chi^2(1) = 3.88$, $p < 0.05$]. These results indicate that the repetitive exposure to somatosensory stimulation during auditory perceptual training can alter the perceptual category boundary as a consequence.

Perceptual Change During the Training

Figure 3 shows the averaged trajectory of the estimated category boundary over the course of training. In order to examine whether the category boundary changed during the training, we applied the LME analysis to the category boundary estimates obtained over the course of the Training

phase. We found that there was no significant interaction between groups (CTL and SOMA) and blocks [$\chi^2(4) = 1.93$, $p > 0.7$], indicating that the pattern of change in the category boundary was similar for the two groups. In addition, we did not find a difference across blocks [$\chi^2(4) = 8.70$, $p > 0.06$], indicating that there was no change in category boundary over the course of the training. There was a significant overall difference between groups [$\chi^2(1) = 5.19$, $p < 0.03$], indicating that the mean value for the category boundary during training in the somatosensory condition was different than that in the control condition. A one-sample t -test using the data averaged across blocks showed that values were significantly different from zero in the SOMA condition [$t(14) = -2.83$, $p < 0.02$], but not in the CTL condition [$t(14) = 1.03$, $p > 0.3$]. This indicates that participants' perception in the SOMA condition shifted in the direction of /ɛ/ during the training phase. This change was not induced in the control condition. The results suggest that there were no temporal changes over the course of training in either group, while somatosensory stimulation induced an overall shift in perception in the experimental condition.

Persistence of Category Boundary Shift

Although we had limited data to evaluate, we assessed whether the perceptual aftereffect persists following training. We first compared category boundary estimates between the first two (1st set) and the last two blocks (2nd set) of the Aftereffect phase using the LME analysis. There were no significant differences between 1st and 2nd sets [$\chi^2(1) = 0.34$, $p > 0.5$]. *Post hoc* tests conducted for the individual conditions found no difference between sets for SOMA ($p > 0.7$) and CTL ($p > 0.6$), respectively. There was a significant interaction between groups and sets [$\chi^2(2) = 11.62$, $p < 0.01$]. *Post hoc* tests indicated a significant difference between SOMA and CTL in the 1st set ($p < 0.05$), and a marginal difference in the 2nd set ($p = 0.073$). There

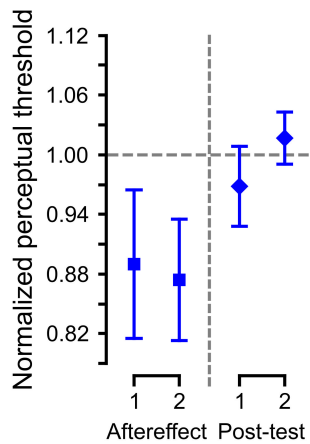


FIGURE 4 | Normalized category boundary in the Aftereffect phase and Post-test (1 week later). Error bars represent the standard error across participants.

is also a significant difference between groups [$\chi^2(1) = 5.14$, $p < 0.05$], such that the values for the SOMA group are different than those of the CTL group. Separate one-sample t -tests showed that the overall mean in the Aftereffect phase in the SOMA group was reliably different than zero [$t(14) = -5.26$, $p < 0.01$], whereas this was not the case for the CTL group [$t(14) = -0.85$, $p > 0.4$]. This indicates that the category boundary change following somatosensory stimulation persisted during Aftereffect trials.

We also evaluated if the perceptual change due to paired auditory-somatosensory stimulation persisted one-week later. Since only five participants from SOMA group were tested following the one-week delay, we evaluated the effects using five datasets for these participants (one pre-training set, two following training and two after a 1 week delay). The averaged data with standard errors for each set of the Aftereffect and Post-test trials are shown in **Figure 4**. The LME analysis showed a significant difference between the Aftereffect and Post-test values [$\chi^2(1) = 4.56$, $p < 0.05$], but not between the 1st and 2nd sets [$\chi^2(1) = 0.11$, $p > 0.7$] nor in the interaction [$\chi^2(2) = 4.95$, $p > 0.08$], suggesting that the somatosensory effect was not present one-week later.

DISCUSSION

The present study examined whether repetitive exposure to somatosensory stimulation in a task which was designed to mirror the pairing of auditory and somatosensory stimulation that occurs during production and speech learning, changes the perceptual representation of speech sounds. We evaluated whether the category boundary between / ϵ / and / a / changed from before to after training with somatosensory stimulation. The somatosensory stimulation involved facial skin deformation in an upward direction. In previous work using a simple perceptual classification task

(Ito et al., 2009), this manipulation was found to change the perception of speech sound toward / ϵ / when presented with the speech stimuli during training. We found instead that the category boundary between / ϵ / and / a / was in fact shifted toward / ϵ /, that is, participants perceived / a / more than / ϵ / after training. Although a relatively small number of participants was available for a subsequent post-training test, the shift in the perceptual boundary did not appear to be present 1 week later. The results nevertheless suggest that repetitive exposure to somatosensory inputs associated with facial skin deformation is capable of changing the perceptual representation of speech sounds.

The results of the present study are in line with previous work showing that facial skin deformation changes the perception of speech sounds in on-line testing (Ito et al., 2009; Trudeau-Fisette et al., 2019; Ogane et al., 2020). Repetitive exposure to somatosensory stimulation during speech motor learning may account for the contribution of somatosensation to speech perception (Ohashi and Ito, 2019). The current results are consistent with this hypothesis. Paired auditory-somatosensory input during training, alters subsequent auditory perceptual judgments, suggesting a contribution of somatosensory exposure to speech perception and the presence of a link between speech production and perception.

As mentioned in the Introduction, the category boundary between vowels can be changed when we are repeatedly exposed to one of two vowels, a phenomenon in the speech perception literature known as selective adaptation. Eimas and Corbit (1973) originally showed that the category boundary between / ba / and / pa / was shifted toward / ba /, that is, the participants perceived / pa / more than / ba / after the training with repetitive exposure of / ba / . The pattern is similar to that of the current finding in which the category boundary shifted toward / ϵ / when repetitive somatosensory stimulation, which has been previously shown to modify the perceived speech sound toward / ϵ /, was applied. A possible mechanism, originally proposed by Eimas and Corbit (1973) is fatigue of a linguistic feature detector as a result of repetitive exposure to the corresponding speech sounds. Kleinschmidt and Jaeger (2016) proposed another possible explanation associated with distributional learning. Although the current results cannot address this debate directly, the current somatosensory effect would fit with either account of selective adaptation. Specifically, in the control condition, we present all values on the speech-sound continuum an equal number of times. As a result, there is no effect on the category boundary, presumably because the entire speech sound representation is affected equally. Both linguistic feature detector and learned distribution accounts would predict a similar result under these conditions. Somatosensory stimulation in the present study serves to modify the perceived speech sound toward / ϵ / . Both feature detection fatigue for / ϵ / and modification of the stimulation distribution would predict this effect which in turn, may be reflected as a change in the category boundary.

Selective adaptation in speech perception is considered to be an auditory phenomenon when the presented sounds

are unambiguous. Previous studies using the McGurk effect (McGurk and MacDonald, 1976) showed that selective adaptation to auditory inputs was induced even when the sound was perceived differently as a result of incongruent visual stimulation (Roberts and Summerfield, 1981; Saldaña and Rosenblum, 1994; Dias et al., 2016). While selective adaptation is observed in visual speech perception (Baart and Vroomen, 2010), Dias et al. (2016) suggested that visual information may not contribute to selective adaptation in the McGurk effect. In the case of the present study, since training with somatosensory stimulation was found to induce a change in the auditory category boundary, the interaction mechanism may be different than in auditory-visual speech perception. This would be consistent with a previous study which found that simultaneous somatosensory and visual stimulation in speech perception did not interact with one other in terms of the behavioral response (Ito et al., 2021). Since somatosensory inputs to speech sounds affect the N1 peak in the auditory ERP (Ito et al., 2014), which is considered to be associated with the initial extraction of vowel related information (Näätänen and Picton, 1987), somatosensory inputs may affect the auditory processing of speech sounds at a lower level of vowel processing. However, somatosensory inputs also affect word segmentation in lexical decisions (Ogane et al., 2020). One future direction is a direct test of the idea that somatosensory stimulation may affect visual speech perception. Baart and Vroomen (2010) showed adaptation in visual speech perception of ambiguous lipread tokens after the exposure to an incongruent sound. Somatosensory stimulation may work in a similar fashion by providing information which disambiguates visual stimuli instead of sounds.

It is important to know how long the training effect lasts. The duration of training phase was limited and as a result, this type of sensory adaptation may not last for a long time. In the case of speech motor learning using altered auditory feedback, the post-training effects on adaptation gradually decrease over the course of the following 100 trials (Purcell and Munhall, 2006; Villacorta et al., 2007). The motion aftereffects described in the Introduction persist for several seconds to minutes. Although it is unknown yet how long selective adaptation lasts, this effect may only persist for a short period, as is the case with sensory adaptation in other modalities. Since the effect of somatosensory training was essentially absent one-week later in the limited number of participants that were tested, the current persistence of a somatosensory aftereffect on speech perception may be similar to other sensory aftereffects. In future investigations, it would be desirable to evaluate shorter periods after training, such as 1 h later, rather than one-week. These types of adaptation including selective adaptation are induced when transient stimulation is presented, and hence when the additional stimulation is removed, particularly after brief periods of training, it is difficult to maintain the adapted perception without receiving additional stimulation. Since this additional stimulation does not exist outside of the laboratory, it may limit the use of the current procedure for speech training or rehabilitation. Nevertheless, the current finding is in line with the more general idea that receiving specific paired

of auditory-somatosensory inputs, such as occurs over long periods of time during speech motor training, may underlie a durable contribution of somatosensory inputs to the speech perceptual representation.

In previous work using the same speech sound continuum, in which skin stretch trials were interleaved with no-stretch trials, a change in perception of speech sound toward /ε/ was observed (Ito et al., 2009). In contrast, in the present study, multiple blocks of 100 trials with skin stretch were used. The repetitive pairing of auditory-somatosensory stimulation may have produced a quite different perceptual effect that favored “selective adaptation.” As a result, participants might have perceived /a/ more than /ε/ even in the first 100 trial block.

A potential technical limitation of the present study is that perceptual boundary between speech sounds could not be estimated over a smaller number of trials. While the MCS provides a reasonable estimate of the perceptual boundary, it requires a relatively large number of trials. In the present study, we used 100 trials, and hence, only five estimates of the boundary value were obtained over the course of the current training. The procedure may thus lack the sensitivity needed to correctly capture any changes which might occur. We used a maximum likelihood procedure before and after training as an alternative to improve the possibility of detecting changes over a shorter period of time. However, this method still needs more than ten trials (17 trials in the current case) and requires that participants listen to sounds near to their perceptual boundary rather than over the entire sound continuum, which is the case with the MCS. Due to this technical limitation, it is difficult to characterize perceptual behavior over the course of training. Further investigation is required to better understand the time-course of the current adaptation mechanism.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comité d’Ethique pour les Recherches non Interventionnelles, Comité d’Ethique pour la Recherche, Grenoble Alpes. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TI contributed to the conception and design of the study. TI and RO collected the data, wrote the first draft of the manuscript, involved in subsequent drafts of the manuscript. RO organized the database, performed the statistical analysis (all under TI’s guidance), and produced the figures. Both authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC017439.

REFERENCES

- Baart, M., and Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. *Neurosci. Lett.* 471, 100-x-103. doi: 10.1016/j.neulet.2010.01.019
- Beauchamp, M. S., Yasar, N. E., Frye, R. E., and Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *Neuroimage* 41, 1011–1020. doi: 10.1016/j.neuroimage.2008.03.015
- Bertelson, P., Vroomen, J., and De Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychol. Sci.* 14, 592–597. doi: 10.1046/j.0956-7976.2003.psci_1470.x
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385. doi: 10.1016/j.cub.2009.01.017
- Dias, J. W., Cook, T. C., and Rosenblum, L. D. (2016). Influences of selective adaptation on perception of audiovisual speech. *J. Phon.* 56, 75–84. doi: 10.1016/j.wocn.2016.02.004
- Eimas, P. D., and Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cogn. Psychol.* 4, 99–109. doi: 10.1016/0010-0285(73)90006-6
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *J. Phon.* 14, 3–28. doi: 10.1016/S0095-4470(19)30607-2
- Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. (2000). Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res.* 10, 77–83. doi: 10.1016/S0926-6410(00)00024-0
- Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502–504. doi: 10.1038/nature08572
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363. doi: 10.1002/bimj.200810425
- Ito, T., and Gomi, H. (2007). Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in speech. *Neuroreport* 18, 907–910. doi: 10.1097/WNR.0b013e32810f2dfb
- Ito, T., Gracco, V. L., and Ostry, D. J. (2014). Temporal factors affecting somatosensory–auditory interactions in speech processing. *Front. Psychol.* 5:1198. doi: 10.3389/fpsyg.2014.01198
- Ito, T., Ohashi, H., and Gracco, V. L. (2021). Somatosensory contribution to audiovisual speech processing. *Cortex* 143, 195–204. doi: 10.1016/j.cortex.2021.07.013
- Ito, T., and Ostry, D. J. (2010). Somatosensory contribution to motor learning due to facial skin deformation. *J. Neurophysiol.* 104, 1230–1238. doi: 10.1152/jn.00199.2010
- Ito, T., Tiede, M., and Ostry, D. J. (2009). Somatosensory function in speech perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1245–1248. doi: 10.1073/pnas.0810063106
- Johansson, R. S., Trulsson, M., Olsson, K. Å, and Abbs, J. H. (1988). Mechanoreceptive afferent activity in the infraorbital nerve in man during speech and chewing movements. *Exp. Brain Res.* 72, 209–214. doi: 10.1007/BF00248519
- Jones, B. C., Feinberg, D. R., Bestelmeyer, P. E. G., DeBruine, L. M., and Little, A. C. (2010). Adaptation to different mouth shapes influences visual perception of ambiguous lip speech. *Psychon. Bull. Rev.* 17, 522–528. doi: 10.3758/PBR.17.4.522
- Kleinschmidt, D. F., and Jaeger, T. F. (2016). Re-examining selective adaptation: fatiguing feature detectors, or distributional learning? *Psychon. Bull. Rev.* 23, 678–691. doi: 10.3758/s13423-015-0943-z
- Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *J. Neurophysiol.* 97, 3155–3164. doi: 10.1152/jn.00086.2007
- Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., and Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *J. Neurosci.* 34, 10339–10346. doi: 10.1523/JNEUROSCI.0108-14.2014
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461.
- Mather, G., Verstraten, F., and Anstis, S. (1998). *The Motion Aftereffect: A Modern Perspective*. Cambridge, MA: A Bradford Book.
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696. doi: 10.1016/j.cub.2007.08.064
- Mochida, T., Kimura, T., Hiroya, S., Kitagawa, N., Gomi, H., and Kondo, T. (2013). Speech Misperception: Speaking and Seeing Interfere Differently with Hearing. *PLoS One* 8:e68619. doi: 10.1371/journal.pone.0068619
- Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x
- Nasir, S. M., and Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20470–20475. doi: 10.1073/pnas.0907032106
- Ogane, R., Schwartz, J.-L., and Ito, T. (2020). Orofacial somatosensory inputs modulate word segmentation in lexical decision. *Cognition* 197:104163. doi: 10.1016/j.cognition.2019.104163
- Ohashi, H., and Ito, T. (2019). Recalibration of auditory perception of speech due to orofacial somatosensory inputs during speech motor adaptation. *J. Neurophysiol.* 122, 2076–2084. doi: 10.1152/jn.00028.2019
- Pinheiro, J., Bates, D., and R Core Team. (2022). *nlme: Linear and nonlinear mixed effects models*. Available online at: <https://svn.r-project.org/R-packages/trunk/nlme/>.
- Purcell, D. W., and Munhall, K. G. (2006). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J. Acoust. Soc. Am.* 120, 966–977. doi: 10.1121/1.2217714
- Roberts, M., and Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Percept Psychophys.* 30, 309–314. doi: 10.3758/bf03206144
- Saldaña, H. M., and Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *J. Acoust. Soc. Am.* 95, 3658–3661. doi: 10.1121/1.409935
- Sams, M., Möttönen, R., and Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Res. Cogn. Brain Res.* 23, 429–435. doi: 10.1016/j.cogbrainres.2004.11.006
- Sato, M., Troille, E., Ménard, L., Cathiard, M.-A., and Gracco, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Exp. Brain Res.* 227, 275–288. doi: 10.1007/s00221-013-3510-8
- Schwartz, J. L., Basirat, A., Menard, L., and Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): a perceptuo-motor theory of speech perception. *J. Neurolinguist* 25, 336–354. doi: 10.1016/J.Jneuroling.2009.12.004
- Shen, Y., and Richards, V. M. (2012). A maximum-likelihood procedure for estimating psychometric functions: thresholds, slopes, and lapses of attention. *J. Acoust. Soc. Am.* 132, 957–967. doi: 10.1121/1.4733540
- Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *J. Acoust. Soc. Am.* 125, 1103–1113. doi: 10.1121/1.3058638
- Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* 25, 76–89. doi: 10.1016/j.neuroimage.2004.11.006
- Sumby, W. H., and Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- Trudeau-Fisette, P., Ito, T., and Ménard, L. (2019). Auditory and Somatosensory Interaction in Speech Perception in Children and

- Adults. *Front. Hum. Neurosci.* 13:344. doi: 10.3389/fnhum.2019.00344
- Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J. Acoust. Soc. Am.* 122, 2306–2319. doi: 10.1121/1.2773966
- Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702. doi: 10.1038/nn1263

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ito and Ogane. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.