HHS Public Access

Author manuscript

Audit Percept Cogn. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:

Audit Percept Cogn. 2023 ; 6(1-2): 97–107. doi:10.1080/25742442.2022.2134674.

Correlation between the effect of orofacial somatosensory inputs in speech perception and speech production performance

Monica Ashokumar[a,*], **Clément Guichet**[a], **Jean-Luc Schwartz**[a], **Takayuki Ito**[a,b]

[a]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

[b]Haskins Laboratories, New Haven, USA

Abstract

Introduction: Orofacial somatosensory inputs modify the perception of speech sounds. Such auditory-somatosensory integration likely develops alongside speech production acquisition. We examined whether the somatosensory effect in speech perception varies depending on individual characteristics of speech production.

Methods: The somatosensory effect in speech perception was assessed by changes in category boundary between /e/ and /ø/ in a vowel identification test resulting from somatosensory stimulation providing facial skin deformation in the rearward direction corresponding to articulatory movement for /e/ applied together with the auditory input. Speech production performance was quantified by the acoustic distances between the average first, second and third formants of /e/ and /ø/ utterances recorded in a separate test.

Results: The category boundary between /e/ and /ø/ was significantly shifted towards /ø/ due to the somatosensory stimulation which is consistent with previous research. The amplitude of the category boundary shift was significantly correlated with the acoustic distance between the mean second – and marginally third – formants of /e/ and /ø/ productions, with no correlation with the first formant distance.

Discussion: Greater acoustic distances can be related to larger contrasts between the articulatory targets of vowels in speech production. These results suggest that the somatosensory effect in speech perception can be linked to speech production performance.

Keywords

Auditory-somatosensory integration; production-perception link; somatosensory feedback; vowel categorization

*Monica Ashokumar monica.ashokumar@grenoble-inp.fr; GIPSA-lab, CNRS, 11 Rue des Mathématiques – BP 46, 38402 Saint Martin d'Hères Cedex, France.

Declaration of Interest

No potential conflict of interest was reported by the authors.

Introduction

Multisensory integration in speech perception has been mostly investigated between the auditory and visual modalities (Benoît et al., 1994; Erber, 1969; Grant & Seitz, 2000; Fort et al., 2010), as wonderfully demonstrated with the well-known McGurk effect (McGurk & MacDonald, 1976). Still, in addition to the visual modality, the somatosensory system also intervenes in speech perception. Indeed, Ito et al (2009) showed that an orofacial somatosensory stimulation associated with facial skin deformation changes the classification of a speech sound between /ɛ/ and /a/ systematically depending on the direction and temporal pattern of the stimulation. This effect was consistently induced both for adults and children (Trudeau-Fisette et al., 2019). Orofacial somatosensory stimulation can also modify the lexical perception of a given sequence involving ambiguous word segmentation, depending on the timing of the stimulation relative to the timing of the corresponding articulatory movement (Ogane et al., 2020).

In these studies, the underlying interpretation could be related to the findings that the somatosensory inputs associated with facial skin deformation provide kinaesthetic information on speech articulatory movements. Indeed, microneurographic studies showed that facial skin deformation by orofacial movements including speech gestures activates cutaneous mechanoreceptors in the orofacial region (Johansson et al., 1988; Nordin & Thomander, 1989). Moreover, behavioural studies showed that external stimulation associated with facial skin deformation induces speech-related compensatory reflex during speaking (Ito & Gomi, 2007) and adaptive change over the course of speech training (Ito and Ostry 2010). These data support the idea that the effect of the somatosensory inputs in speech perception could be related to their role in speech production. This is in line with the role of somatosensory interactions in the recalibration of speech perception by speech motor learning. In fact, Ohashi and Ito (2019) specifically showed that somatosensory inputs applied during a speech motor learning task induced a recalibration of speech perception. Perceptual training with somatosensory inputs also changed the perceptual boundary in a vowel categorization task (Ito and Ogane 2022). Altogether, these studies are in line with the assumption that the role of somatosensory inputs in speech perception is related to a functional link between speech production and speech perception processes.

A link between speech production and perception has been proposed and investigated for more than 50 years within the framework of the Motor Theory of Speech Perception (Liberman et al., 1967) and the Direct Realist Theory of Speech Perception (Fowler, 1986). More recently, the Perception-for-Action-Control Theory (PACT) (Schwartz et al., 2012) suggests that a speech unit is neither purely auditory nor motor but rather perceptuo-motor. Thus, procedural knowledge from speech production together with sensory information from the auditory, visual and somatosensory systems which shape the articulatory gestures, both contribute to the makeup of phonetic representations and, ultimately, of the speech perception process. The PACT framework accounts for the perceptual modulation induced by speech motor learning using a Bayesian model (Patri et al., 2018). The principled link between speech perception and speech production at work in PACT allows to hypothesise that the auditory-somatosensory integration mechanism involved in speech perception could be calibrated based on performance in speech production, and hence depend on the acquisition of speech production mastery for each individual.

The principle of a production-perception link has also been directly investigated by comparing speech production and speech perception abilities. Indeed, it has been repeatedly shown that better auditory acuity in a speech sound discrimination task is associated with larger articulatory and acoustic distances between speech targets and hence more precise targets in speech production (Franken et al., 2017; Perkell et al., 2004). Chao et al. (2019) also demonstrated that the perceptual category boundary between two target vowels was closer to the category with the smaller variability in production, and hence that the boundary between the two vowel categories in production, determined according to the variability of utterances for each category, was strongly correlated across individuals with the perceptual category boundary between the two vowels. Furthermore, Ghosh et al. (2010) showed that variations in auditory and somatosensory discrimination were also correlated with the magnitude of contrast between the sibilants /s/ and /ʃ/. All this suggests that speech production and speech perception abilities are closely linked and shaped in relation with individual variability.

It is generally considered that the production-perception link is settled and/or shaped through development. Speech production abilities are different between adults and children. For example, adults show less variability and more acoustic contrast between vowels in speech production (e.g. Vorperian and Kent 2007). While sensorimotor influences in the processing of speech perception can be observed in infancy (DePaulis et al. 2011, Bruderer et al. 2015), Vilain et al., (2019) specifically showed that the emergence of babbling in infants is important for the extraction of vowel-independent plosive representations, and Kuhl et al. (2014) showed cortical maturation of the perceptuo-motor link in relation to speech production development. In consequence, since the somatosensory role in speech perception is supposed to be related to speech production processes, this role should evolve along with the development of speech production. Indeed, Trudeau-Fisette et al (2019) showed that the size of the somatosensory effect in vowel perception was larger in adults than in children. This is also in line with the view that the development of multisensory integration depends upon the maturation of individual sensory systems (Gori, 2015). Considering that differences in developmental trajectories result in individual variability in speech production, our assumption is that the variability in speech production might also be related to the variability in the magnitude of the somatosensory effect in speech perception.

The current study explores this assumption. The experiment consists of a within-subjects procedure, in which participants are involved in both a speech perception task with somatosensory stimulation and a speech production task. The somatosensory effect in speech perception was quantified by the amount of category shift in a vowel identification test between the front mid-high French vowels /e/ (unrounded) and /ø/ (rounded), as performed in Trudeau-Fisette et al. (2019). For characterizing speech production performance, we recorded the corresponding vowels /e/ and /ø/ from the same participants, and evaluated the acoustic distance in the first, second and the third formant frequencies (F1, F2 and F3) between those vowels. Then, we carried out correlation analyses between acoustic distances in speech production and the amplitude of the somatosensory effect in speech perception. This analysis was applied in F1, F2 and F3 respectively. As known from previous studies, we expected that F2 would predominantly represent the acoustical contrast between /e/ and /ø/, while F1 should be stable from one category to the other (Ménard et al., 2008). F3 may also separately represent the articulatory lip rounding contrast for high front vowels (Lindblom & Sundberg, 1971; Schwartz et al., 1993). Therefore, we expected to observe a correlation with F2 and possibly F3 but not with F1.

Methods

Participants

Nineteen native speakers of French participated in this study. The participants had no neurological deficits in hearing and speaking. All the participants signed the consent form approved by the local ethical committee of the Université Grenoble Alpes [Comité d'Ethique pour la Recherche, Grenoble Alpes (CERGA-Avis-2021-8)].

Experimental Procedure

All participants took part in both a speech perception test with somatosensory stimulation and a speech production test. These two tests were carried out in a single session. The perception test with somatosensory stimulation was followed by the production test. For the speech perception test with somatosensory stimulation, we applied the vowel identification test used in Trudeau-Fisette et al., (2019), in which participants were asked to identify the presented vowel stimuli as /e/ or /ø/. The auditory stimuli were the ones produced in Trudeau-Fisette et al. (2019). Those are isolated vowel sounds with 300 ms duration and consisted of 8 members of a synthesised /e/-/ø/ continuum (stimuli no. 1 to 8 in Trudeau-Fisette et al. (2019)). The formant values of the speech stimuli and details of the synthesis procedure can be found in Trudeau-Fisette et al. (2019). Briefly, the continuum was synthesized according to the procedure in Ménard & Boe (2004) using an articulatory synthesizer. In this continuum, F2, F3 and F4 were equally shifted in steps of 30, 40 and 50 Hz respectively while F1 and F5 were kept constant respectively at 364 Hz and 4000 Hz. The values of F2, F3 and F4 were respectively at 1922, 2509, and 3550 Hz for /e/ at one edge of the continuum, and at 1712, 2229, and 3200 for /ø/ at the other end. Auditory stimuli were presented via a loudspeaker, set in front of the participant, with a comfortable listening level. The somatosensory perturbation associated with facial skin stretch was presented with a small robotic device (Phantom 1.0, 3D Systems; see all details on this system in Ito et al., 2009). The robot has a wire on its arm. Two plastic tabs (2 × 3 cm) from the other end of the wire are attached laterally to the sides of the mouth of the participant with double-sided tape (Figure 1(a)). The system was programmed to produce a sinusoidal temporal pattern with a peak force of 4N, with a 90-ms lead of the somatosensory stimulation onset relative to the auditory stimulus onset (Figure 1(b)). This 90-ms somatosensory lead was chosen because it successfully altered perception of speech sounds in a previous study (Ito et al., 2009).

The somatosensory perturbation was applied in the rearward (posterior) direction based on the assumption that skin deformation along this axis can represent somatosensory inputs corresponding to articulatory movement for production of the lip-spread vowel /e/. We tested two conditions: auditory condition (control) and auditory-somatosensory condition (skin-stretch). These two conditions were alternated every 8 trials, in which the 8 members of the synthesised /e/-/ø/ continuum were presented in random order. In total, 24 blocks of 8 trials were tested.

For assessing speech production performance, we recorded the monosyllabic words “dé” (English “dice”, /de/) and “deux” (English “two”, /dø/) using a picture naming test. The participants were asked to speak aloud the word corresponding to the picture shown on a monitor in a usual speaking manner. Six repetitions were recorded for each word. The current experiment is part of a larger protocol, which is designed for hearing impaired individuals with hearing aids/cochlear implant devices. In this protocol involving a longer recording session, we also carried out a noisy-digit identification test to assess speech perception in noise, together with recordings of speech production for a broad range of vowels. This is why, in order to make the total experimental time acceptable, we recorded only six repetitions in the present vowel production task.

Statistical Analysis

In the perceptual experiment, the probability of /ø/ responses was calculated for each of the 8 stimuli for each participant and each condition. Using these judgement probability values, individual psychometric functions were estimated in each condition (control and skin-stretch), to obtain the point of subjective equality (PSE) at the 50% crossover boundary value. This PSE value represents the category boundary between the two vowels and has been used to assess the somatosensory effect in previous studies (Ito et al. 2009, Trudeau-Fisette et al. 2019). Repeated-measures ANOVA was applied to compare PSE values between the conditions (control and skin-stretch). The amplitude of the somatosensory effect in speech perception was quantified using the difference in PSE between the skin-stretch and control conditions.

F1, F2 and F3 frequencies were extracted using PRAAT (Boersma & Weenink, 2001) in order to evaluate individual variations in production performance. The obtained formants were temporally averaged in a 40 ms time window centered on the middle of the manually detected period of target vowel production (see Figure 3(a)). The speech production index was computed separately for F1, F2 and F3 as the acoustic distance in Hz between mean values of the six /e/ vs. /ø/ utterances.

Correlation analyses with Pearson correlation coefficients between perception and production performance among the 19 participants were then separately computed between the index of somatosensory effect in speech perception and each of the three indices of speech production as characterized respectively by F1, F2 and F3.

Results

The perceptual data successfully replicated the results obtained by Trudeau-Fisette et al. (2019), showing that perception of the /e/-/ø/ contrast was modified by the somatosensory perturbation associated with facial skin deformation. Figure 2(a) shows a representative example of the judgment probability for each of the 8 stimuli and the estimated psychometric function of an individual participant. The estimated psychometric function was shifted towards the right (i.e. towards /ø/) in the skin-stretch condition, indicating that more stimuli were identified as the vowel /e/ in the skin-stretch condition. Figure 2(b) shows the mean PSE values in control and skin-stretch conditions. A one-way repeated-measures ANOVA showed a significant difference between control and skin-stretch conditions ($F(1,18) = 7.42$, $p = 0.01$).

Considering production data, Figure 3(a) represents the temporal pattern of the sound signal for one utterance of the word “dé” and “deux” (top panel) and the corresponding spectrogram (bottom panel). Horizontal solid lines represent the three formants (F1, F2 and F3) extracted using PRAAT. As shown in this representative example, F1 is similar, but F2 and F3 are different between the two vowels. Figure 3(b) represents the formant values averaged across participants, with the corresponding standard error. Notice that the variability in formant values from one utterance to another of the same vowel is rather low for all participants (standard deviations of formant values for each vowel category and each participant and then averaged across participants are around 15 Hz for F1, 50 Hz for F2 and 60 Hz for F3 for both /e/ and /ø/). The acoustic distance between mean values for /e/ and /ø/ stays very small in F1, but large in F2 and F3. As expected, the acoustical contrast between /e/ and /ø/ hence appears as well represented in F2 and in F3.

We then correlated the somatosensory effect in vowel identification (PSE difference) with speech production indices that are acoustic distance in F1, F2 or F3 between the target vowels /e/ and /ø/. Figure 4 represents scatter plots together with the regression line and 95 % confidence interval for each correlation display. While there is no significant correlation with F1 ($r(17) = 0.17$, $p = 0.49$), there is a significant correlation with F2 ($r(17) = 0.48$, $p = 0.03$) and correlation is marginally significant with F3 ($r(17) = 0.41$, $p = 0.08$). This indicates that larger somatosensory effects in speech perception are obtained for participants who have a larger difference between /e/ and /ø/ in F2 and/or F3 in speech production. Since the target vowels are contrasted in F2 and in F3 as shown in Figure 3, these correlation results suggest that the production of the contrast for the corresponding vowels can be related to the somatosensory-auditory interaction in speech perception.

Discussion

In the present study, we examined whether the somatosensory effect in speech perception correlates with individual variability of speech production. We successfully replicated the findings by Trudeau-Fisette et al (2019) showing that the category boundary between /e/ and /ø/ was shifted towards /ø/ when a somatosensory stimulation associated with rearward skin stretch was applied. Importantly, we found a reliable correlation between the size of this somatosensory effect and the acoustic distance between /e/ and /ø/. This correlation was seen in F2 and marginally in F3, but not in F1.

The results from the current study confirm our expectation that the somatosensory effect and, accordingly, the auditory-somatosensory integration mechanism in speech perception might vary in relation with the production performance of participants. They shed light on the way this relation seems to operate, with larger acoustic distances between vowel targets leading to a larger somatosensory effect. A possible interpretation, in line with the previously cited studies showing a link between sensory and motor precision (Franken et al., 2017; Ghosh et al., 2010; Perkell et al., 2004), could be that if /e/ and /ø/ targets are far apart, this should correspond to a large articulatory gesture with a large movement of the face, associated with a large contrast in somatosensory configurations between the targets. Individuals with such a large acoustic distance would hence be more sensitive to the somatosensory contrast between the vowels, and in consequence, display a larger somatosensory effect in perception. The correlation between the somatosensory effect in perception and the acoustic distance between /e/ and /ø/ targets in production is significant for F2 but only marginally for F3, which could be related to the larger acoustic variation between the two vowels in F2 than in F3 (see Fig. 3, showing that the mean distance is 547 Hz in F2 vs. 345 Hz in F3, with also lower standard deviations in F2 than in F3). This is likely related to the fact that F3 is known to be a correlate of lip rounding for high (/i/ vs. /y/) rather than for mid-high (/e/ vs. /ø/) vowels (Lindblom & Sundberg, 1971).

The correlation between speech production performance and the magnitude of the somatosensory effect in speech perception fits well with the assumption that the development of the somatosensory function and of auditory-somatosensory integration in speech perception could be related to the acquisition of speech production. This possibility is also in line with the findings by Trudeau-Fisette et al. (2019) that the somatosensory effect in the perception of speech sounds was larger in adults than in children. Indeed, the not-yet-mature speech production in children would result in having an auditory-somatosensory integration still under development, and hence a smaller somatosensory effect in speech perception for children compared to adults. Interestingly, such developmental differences in auditory-somatosensory integration related to speech production abilities could have important consequences considering hearing impaired individuals who use cochlear implants. Since there is a critical period concerning neural plasticity in the auditory cortex during early childhood (Kral & Sharma, 2012), both hearing and speaking abilities of these individuals are very dependent on age of implantation and usage period of the cochlear implants (Cardon et al., 2012). We therefore anticipate that investigating the somatosensory effect in hearing impaired individuals equipped with cochlear implants could provide further information on such underlying mechanisms, and the relationship between this effect and speech production performance.

Funding

This work was supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Sklowdoska-Curie Grant Agreement No 860755 (Comm4CHILD project) and by National Institute on Deafness and Other Communication Disorders Grant R01-DC017439.

References

Benoît C, Mohamadi T, & Kandel S (1994). Effects of Phonetic Context on Audio-Visual Intelligibility of French. Journal of Speech, Language, and Hearing Research, 37(5), 1195–1203. 10.1044/jshr.3705.1195

Boersma P, & Weenink D (2001). PRAAT, a system for doing phonetics by computer. Glot International, 5, 341–345.

Bruderer AG, Danielson DK, Kandhadai P, Werker JF (2015) Sensorimotor influences on speech perception in infancy. Proc Natl Acad Sci U S A 112:13531–13536. [PubMed: 26460030]

Cardon G, Campbell J, & Sharma A (2012). Plasticity in the developing auditory cortex: Evidence from children with sensorineural hearing loss and auditory neuropathy spectrum disorder. Journal of the American Academy of Audiology, 23(6), 396–411; quiz 495. 10.3766/jaaa.23.6.3 [PubMed: 22668761]

Chao S-C, Ochoa D, & Daliri A (2019). Production Variability and Categorical Perception of Vowels Are Strongly Linked. Frontiers in Human Neuroscience, 13, 96. 10.3389/fnhum.2019.00096 [PubMed: 30967768]

DePaolis RA, Vihman MM, & Keren-Portnoy T (2011). Do production patterns influence the processing of speech in prelinguistic infants? Infant Behavior and Development, 34(4), 590–601. 10.1016/j.infbeh.2011.06.005 [PubMed: 21774986]

Erber NP (1969). Interaction of audition and vision in the recognition of oral speech stimuli. Journal of Speech and Hearing Research, 12(2), 423–425. 10.1044/jshr.1202.423 [PubMed: 5808871]

Fort M, Spinelli E, Savariaux C, & Kandel S (2010). The word superiority effect in audiovisual speech perception. Speech Communication, 52, 525–532. 10.1016/j.specom.2010.02.005

Fowler CA (1986). An event approach to the study of speech perception from a direct–realist perspective. Journal of Phonetics, 14(1), 3–28. 10.1016/S0095-4470(19)30607-2

Franken MK, Acheson DJ, McQueen JM, Eisner F, & Hagoort P (2017). Individual variability as a window on production-perception interactions in speech motor control. The Journal of the Acoustical Society of America, 142(4), 2007–2018. 10.1121/1.5006899 [PubMed: 29092613]

Ghosh SS, Matthies ML, Maas E, Hanson A, Tiede M, Ménard L, Guenther FH, Lane H, & Perkell JS (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. The Journal of the Acoustical Society of America, 128(5), 3079–3087. 10.1121/1.3493430 [PubMed: 21110603]

Gori M (2015). Multisensory Integration and Calibration in Children and Adults with and without Sensory and Motor Disabilities. Multisensory Research, 28(1–2), 71–99. 10.1163/22134808-00002478 [PubMed: 26152053]

Grant KW, & Seitz PF (2000). The use of visible speech cues for improving auditory detection of spoken sentences. The Journal of the Acoustical Society of America, 108(3 Pt 1), 1197–1208. 10.1121/1.1288668 [PubMed: 11008820]

Ito T, & Gomi H (2007). Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in speech. Neuroreport, 18(9), 907–910. 10.1097/WNR.0b013e32810f2dfb [PubMed: 17515799]

Ito T, & Ogane R (2022). Repetitive Exposure to Orofacial Somatosensory Inputs in Speech Perceptual Training Modulates Vowel Categorization in Speech Perception. Frontiers in Psychology, 13, 839087. 10.3389/fpsyg.2022.839087 [PubMed: 35558689]

Ito T, & Ostry DJ (2010). Somatosensory Contribution to Motor Learning Due to Facial Skin Deformation. Journal of Neurophysiology, 104(3), 1230–1238. 10.1152/jn.00199.2010 [PubMed: 20592121]

Ito T, Tiede M, & Ostry DJ (2009). Somatosensory function in speech perception. Proceedings of the National Academy of Sciences, 106(4), 1245–1248. 10.1073/pnas.0810063106

Johansson RS, Trulsson M, Olsson KÅ, & Abbs JH (1988). Mechanoreceptive afferent activity in the infraorbital nerve in man during speech and chewing movements. Experimental Brain Research, 72(1), 209–214. 10.1007/BF00248519 [PubMed: 3169190]

Kral A, & Sharma A (2012). Developmental neuroplasticity after cochlear implantation. Trends in Neurosciences, 35(2), 111–122. 10.1016/j.tins.2011.09.004 [PubMed: 22104561]

Kuhl PK, Ramírez RR, Bosseler A, Lin J-FL, & Imada T (2014). Infants' brain responses to speech suggest Analysis by Synthesis. Proceedings of the National Academy of Sciences, 111(31), 11238–11245. 10.1073/pnas.1410963111

Liberman AM, Cooper FS, Shankweiler DP, & Studdert-Kennedy M (1967). Perception of the speech code. Psychological Review, 74(6), 431–461. 10.1037/h0020279 [PubMed: 4170865]

Lindblom BE, & Sundberg JE (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. The Journal of the Acoustical Society of America, 50(4), 1166–1179. 10.1121/1.1912750 [PubMed: 5117649]

Mcgurk H, & Macdonald J (1976). Hearing lips and seeing voices. Nature, 264(5588), 746–748. 10.1038/264746a0 [PubMed: 1012311]

Ménard L, & Boe L (2004). L'emergence du systeme phonologique chez l'enfant: L'apport de la modelisation articulatoire. 10.1353/CJL.2005.0003

Ménard L, Schwartz J-L, & Aubin J (2008). Invariance and variability in the production of the height feature in French vowels. Speech Communication, 50(1), 14–28. 10.1016/j.specom.2007.06.004

Nordin M, & Thomander L (1989). Intrafascicular multi-unit recordings from the human infra-orbital nerve. Acta Physiologica Scandinavica, 135(2), 139–148. 10.1111/j.1748-1716.1989.tb08561.x [PubMed: 2922998]

Ogane R, Schwartz J-L, & Ito T (2020). Orofacial somatosensory inputs modulate word segmentation in lexical decision. Cognition, 197, 104163. 10.1016/j.cognition.2019.104163 [PubMed: 31891832]

Ohashi H, & Ito T (2019). Recalibration of auditory perception of speech due to orofacial somatosensory inputs during speech motor adaptation. Journal of Neurophysiology, 122(5), 2076–2084. 10.1152/jn.00028.2019 [PubMed: 31509469]

Patri J-F, Perrier P, Schwartz J-L, & Diard J (2018). What drives the perceptual change resulting from speech motor adaptation? Evaluation of hypotheses in a Bayesian modeling framework. PLOS Computational Biology, 14(1), e1005942. 10.1371/journal.pcbi.1005942 [PubMed: 29357357]

Perkell JS, Guenther FH, Lane H, Matthies ML, Stockmann E, Tiede M, & Zandipour M (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. The Journal of the Acoustical Society of America, 116(4), 2338–2344. 10.1121/1.1787524 [PubMed: 15532664]

Schwartz J-L, Basirat A, Ménard L, & Sato M (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. Journal of Neurolinguistics, 25(5), 336–354. 10.1016/j.jneuroling.2009.12.004

Schwartz J-L, Beautemps D, Abry C, & Escudier P (1993). Inter-individual and cross-linguistic strategies for the production of the [i] vs. [y] contrast. Journal of Phonetics, 21(4), 411–425.

Trudeau-Fisette P, Ito T, & Ménard L (2019). Auditory and Somatosensory Interaction in Speech Perception in Children and Adults. Frontiers in Human Neuroscience, 13, 344. 10.3389/fnhum.2019.00344 [PubMed: 31636554]

Vilain A, Dole M, Lœvenbruck H, Pascalis O, & Schwartz J-L (2019). The role of production abilities in the perception of consonant category in infants. Developmental Science, 22(6), e12830. 10.1111/desc.12830 [PubMed: 30908771]

Vorperian HK, Kent RD (2007) Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. J Speech Lang Hear Res 50:1510–1545. [PubMed: 18055771]

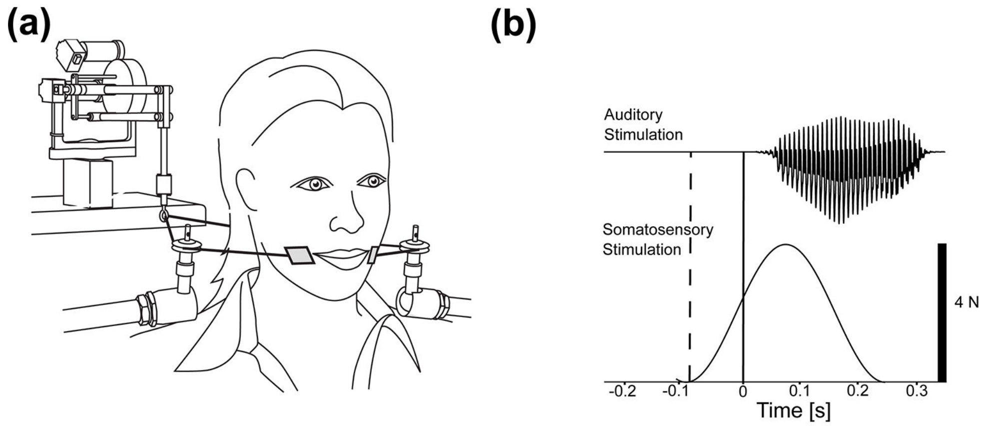

Figure 1.

(a) Experimental setup for vowel identification test with somatosensory stimulation. Reproduced with permission from (Ito & Ostry, 2010). (b) Temporal relationship between the somatosensory stimulation (onset, dotted line) and the auditory stimulation (onset, plain line).

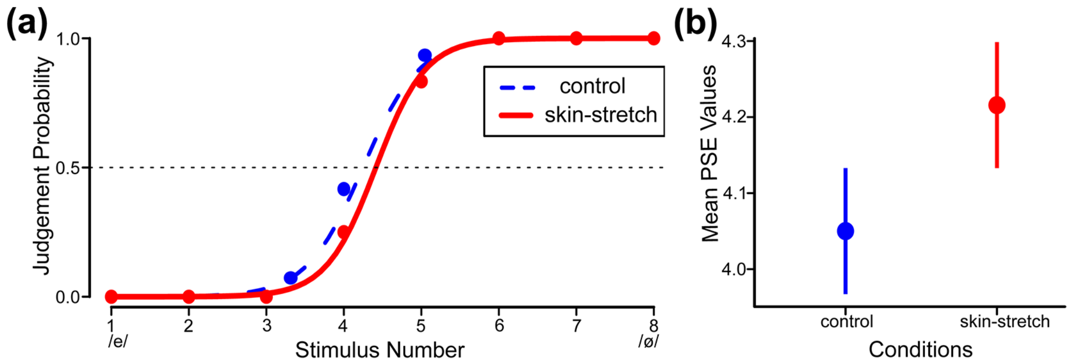

Figure 2.

(a) Judgment probabilities and estimated psychometric function for a representative participant. The horizontal dotted line shows the 50th percentile of the judgment probability. The point where it crosses the estimated psychometric function is the PSE which represents the category boundary between the two vowel targets. (b) Mean PSE values in control and skin-stretch conditions. Error bars show the standard error.

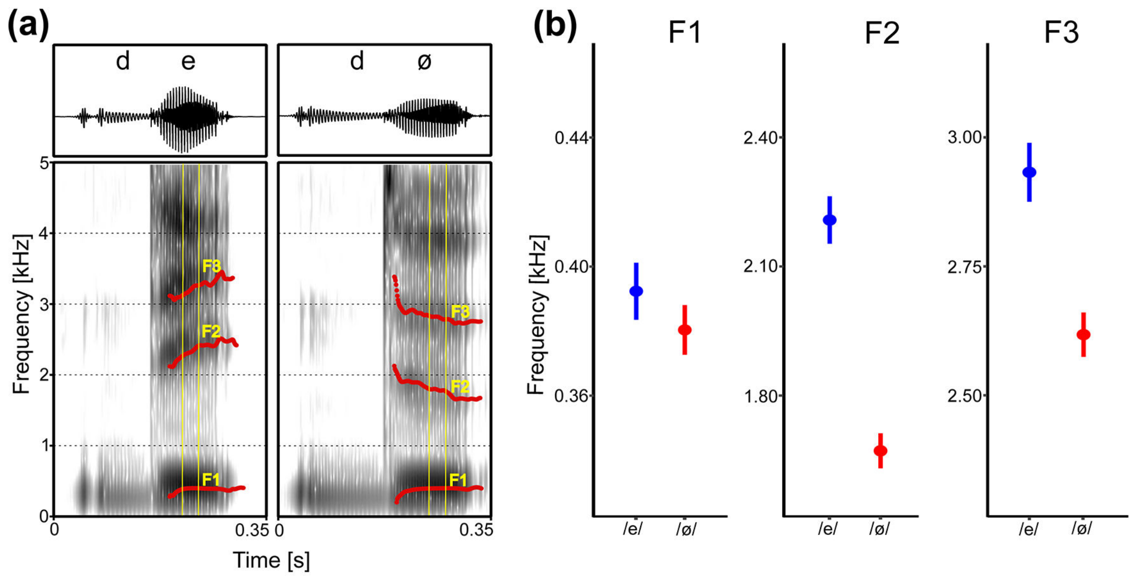

Figure 3.

(a) Temporal pattern of target words ‘dé’ and ‘deux’ (top panel) and corresponding spectrograms (bottom panel). The trajectories of the three formants (F1, F2 and F3) extracted using PRAAT are displayed by thick solid lines. The 40ms time window used for formant measurement is temporally set at the middle of the vowel and marked by the two vertical thin solid lines. (b) Averaged formant values across participants. Error bars show standard error.

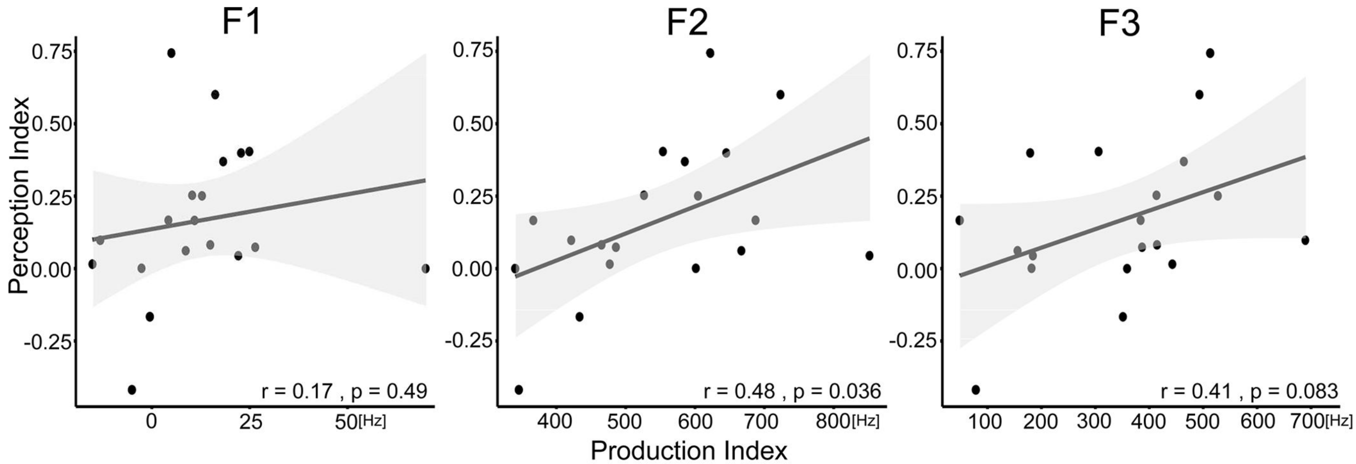

Figure 4.

Scatterplots of perception index (PSE differences) against production index (differences between /e/ and /ø/ in F1, F2 and F3). Solid lines represent the regression line and the shaded area represents 95% confidence intervals.