Orofacial somatosensory inputs modulate word segmentation in lexical decision

Rintaro Ogane, Jean-Luc Schwartz, Takayuki Ito

Title: Orofacial Somatosensory Inputs Modulate Word Segmentation in Lexical Decision

Author names: Rintaro Ogane[1], Jean-Luc Schwartz[1], Takayuki Ito[1,2]

Affiliations:

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

[2]Haskins Laboratories, 300 George Street, New Haven, CT 06511, USA

Word count (Abstract through Conclusions): 6403 words

Corresponding author:

Rintaro Ogane, PhD.

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab

11 rue des Mathématiques, Grenoble Campus, BP46, F-38402 SAINT MARTIN D'HERES Cedex, France

Tel: +33 (0)4 76 82 71 31

Email: rintaro.ogane@gipsa-lab.grenoble-inp.fr

Abstract

There is accumulating evidence that articulatory/motor knowledge plays a role in phonetic processing, such as the recent finding that orofacial somatosensory inputs may influence phoneme categorization. We here show that somatosensory inputs also contribute at a higher level of the speech perception chain, that is, in the context of word segmentation and lexical decision. We carried out an auditory identification test using a set of French phrases consisting of a definite article “la” followed by a noun, which may be segmented differently according to the placement of accents within the phrase. Somatosensory stimulation was applied to the facial skin at various positions within the acoustic utterances corresponding to these phrases, which had been recorded with neutral accent, that is, with all syllables given similar emphasis. We found that lexical decisions reflecting word segmentation were significantly and systematically biased depending on the timing of somatosensory stimulation. This bias was not induced when somatosensory stimulation was applied to the skin other than on the face. These results provide evidence that the orofacial somatosensory system contributes to lexical perception in situations that would be disambiguated by different articulatory movements, and suggests that articulatory/motor knowledge might be involved in speech segmentation.

Keywords

lexical access, articulatory knowledge, speech perception, speech production, perceptuo-motor interaction, multisensory interactions.

1. Introduction

1.1 Perceptuo-motor relationships and the role of somatosensory information in phonetic decoding

A long-standing question about speech perception concerns the potential role of articulatory knowledge in the phonetic decoding process. Coarticulatory phenomena classically modify the acoustic content of a given phonemic unit, which led to the development of the Motor Theory of Speech Perception, arguing that speech decoding is based on the recovery of the motor cause of speech stimuli, and that articulatory/motor representations provide the basis of speech communication (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). Listening to speech sounds activates cortical areas related to speech production in the motor and premotor cortex (e.g. Fadiga et al., 2002; Grabski et al., 2013; Pulvermüller et al., 2006; Tremblay & Small, 2011; Watkins, Strafella, & Paus, 2003; Wilson et al., 2004). A number of behavioral studies show that articulatory movements preceding or accompanying the presentation of auditory stimuli modify speech perception, by e.g. motor stimulation (Sato et al., 2011) or articulatory suppression (Stokes, Venezia, & Hickok, 2019). Articulatory training by imitation appears to improve the auditory comprehension of an unfamiliar accent (Adank, Hagoort, & Bekkering, 2010) or a dysarthric speaker (Borrie & Schäfer, 2015) and training articulation with altered auditory feedback changes further perception of speech sounds (Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Shiller, Sato, Gracco, & Baum, 2009). Importantly however, the effects of articulation on perception are generally small and mostly obtained in configurations for which auditory decoding is made difficult because of noise, natural or induced degradation or stimulus ambiguity (see e.g. D'Ausilio, Jarmolowska, Busan, Bufalari, & Craighero, 2011; Stokes et al., 2019).

Somatosensory information associated to speech articulation is likely to play an important role in this process. The orofacial somatosensory system differs from the limb system and other body parts in terms of proprioceptive function since muscle proprioceptors, which play a predominant role in proprioception, have not been found in the orofacial muscles besides the jaw closing muscles (Stål, Eriksson, Eriksson, & Thornell, 1990). Given that the facial skin is deformed in orofacial movements including speaking (Connor & Abbs, 1998), cutaneous mechanoreceptors in the facial skin can play a role as alternative sources of proprioceptive information. Previous neural recording observations confirmed that cutaneous

mechanoreceptors lateral to the oral angle are activated in jaw motion (Johansson, Trulsson, Olsson, & Abbs, 1988; Nordin & Hagbarth, 1989). This idea has also been demonstrated in somatosensory perturbation studies applying facial skin deformation externally. Ito & Gomi (2007) showed that downward skin stretches laterally to the oral angle induced compensatory reflex response in the upper lip related to jaw downward movements. Stretching the skin backwards also induced adaptive movement change in the upper lip for utterances requiring lip protrusion (Ito & Ostry, 2010). Accordingly, stretching the facial skin in a specific direction can provide somatosensory information related to lip and jaw articulatory motion, and can be used as an effective tool to investigate the orofacial somatosensory function in the processing of speech sounds.

Indeed, the role of somatosensory inputs arising from the facial skin in speech perception has been displayed by Ito, Tiede, & Ostry (2009). These authors reported that when the facial skin was pulled in the upward direction, an auditory stimulus ambiguous between /head/ and /had/ was identified as /head/ rather than /had/. Their interpretation was that articulatory motion for /head/ and /had/ involves vertical movements of the jaw and tongue, allowing modulations of the perception of speech sounds in this region by applying adequate somatosensory input. This kind of studies suggests a potential role of the somatosensory system in speech perception, in relation with theoretical proposals associating auditory processes and articulatory inferences in multisensory theories of speech perception (Schwartz, Basirat, Ménard, & Sato, 2012; Skipper, Van Wassenhove, Nusbaum, & Small, 2007).

1.2 Assessing the role of the somatosensory system in lexical access for speech perception

Coarticulatory processes not only make the acoustic content of a phonemic unit context-dependent, but may also intervene to blur or enhance the segmentation process, crucial for lexical access (Spinelli, Grimault, Meunier, & Welby, 2010; Spinelli, Welby, & Schaegis, 2007). Since coarticulatory processes are based on articulatory mechanisms related to anticipation and perseveration in gestural dynamics, it is likely that the structure of articulatory motion plays a role in the segmentation as well as the decoding process. Considering the role of the somatosensory system in phonetic decoding, the question that we asked in this study is whether it could also intervene at the level of word segmentation for lexical access. This would provide a hint that perceptuo-motor relationships are more pervasive in speech perception than currently envisioned,

and that they actually structure the processing chain enabling to relate the incoming speech signals with the lexicon in the human brain.

For this aim, we capitalized on the paradigm by Spinelli et al. (2010) on the role of prosodic cues for disambiguation of ambiguous acoustic structures in French. The study tested French phrases consisting of a definite article "la" followed by a noun, which are pronounced in the same way because of "elision" phenomena, e.g. "l'attache", /l#ataʃ/ ["the string" in English] vs. "la tache", /la#taʃ/ ["the stain" in English], "#" indicating the word boundary. The authors found that acoustic prosodic cues (e.g. local F0 increase) enabled to switch the percept from one structure to the other, and suggested that the phrases can be disambiguated and segmented differently according to the placement of the accents in the utterance, in line with articulatory strategies displayed in the production of this kind of material (Spinelli et al., 2007).

Since putting an accent in a phrase or changing the acoustic prosodic cues can be achieved by hyper-articulation (Fougeron, 2001; Spinelli et al., 2007), the cues for word segmentation may be obtained not only from acoustical information, but also from articulatory information provided by other sensory modalities. It has been known for a long time that the visual modality contributes to speech perception, not only for phonetic decoding (e.g. for speech in noise, Erber, 1969; Sumby & Pollack, 1954; or with incongruent auditory and visual inputs, McGurk & MacDonald, 1976) but also in prosodic processing (Dohen, Lœvenbruck, Cathiard, & Schwartz, 2004), lexical access (Fort et al., 2013) and word segmentation (Mitchel & Weiss, 2014; Sell & Kaschak, 2009). A recent study (Strauß, Savariaux, Kandel, & Schwartz, 2015), using the same type of material as Spinelli et al. (2010), confirmed that accentuated visual lip movements at a given position in the phonetic input may attract the perceptual placement of word segmentation, suggesting that visual lip information can play a role similar to acoustic prosody. Given that facial skin deformation has already been shown to provide articulatory information able to modify the phonetic decoding process, it might also contribute to modify the segmentation process before lexical access in the processing of a continuous speech stream.

The present study aims at exploring whether somatosensory inputs associated with facial skin deformation could also intervene in the segmentation process and hence modify lexical decision in French. To test this hypothesis, we carried out an auditory identification test of word segmentation similar to the one by Spinelli et al. (2010) and Strauß et al. (2015), using a specific lexical material in French characteristic of the elision phenomenon introduced previously. We examined how perceptual performance in an auditory identification test was modulated depending on when somatosensory inputs were applied during listening at the target auditory phrases. We speculated that a somatosensory stimulation pulling the facial skin upwards (as in Ito et al., 2009) at a given instant would lead participants to infer the presence of an accent around the corresponding position in time, and that this would modify the result of the segmentation process. We further speculated that a somatosensory stimulation applied elsewhere on the body (here, on the forearm) would be less or not effective. Finally, since multisensory interaction requires adequate matching of the various sources of information between the involved modalities, we reasoned that the vertical facial skin deformation would be more effective for utterances containing vowels realized with vertical articulatory movements of the jaw and tongue (e.g. /a/) than horizontal tongue or lip movements (e.g. /i/ or /o/).

2. Methods

2.1 Participants

Forty native French speakers (mean age = 27.10 years, *SD* = 6.56 years, 11 males, 29 females) participated in the experiment. They had no record of neurophysiological issues with hearing or orofacial sensation. The protocol of this experiment was approved by the Comité d'Ethique pour la Recherche, Grenoble Alpes (CERGA: Avis-2018-12-11-4). All participants signed the corresponding consent form.

2.2 Acoustic material

The acoustic material was directly inspired from Spinelli et al. (2010). It consisted in sequences of a definite article “la” and a noun in French. Because of elision, such sequences can be segmented in two possible phrases with different nouns though with the same phonemic sequence, e.g. “l’attache” vs. “la tache”. One of the possible nouns begins with a vowel (V-onset word, e.g. “attache”) and the other one with a consonant (C-onset word, e.g. “tache”). Each pair of possible phrases can be disambiguated by manipulating prosodic cues, by e.g. hyper-articulating the first or second vowel in the article+noun sequence (Spinelli et al., 2007). Each article+noun sequence was preceded by a carrier phrase “C’est” [“This is” in English] to produce a complete French sentence.

We tested seventeen pairs of French words (Table 1). The auditory stimuli spoken by a native French male speaker were digitally recorded at a sampling frequency of 44.1 kHz. In order to minimize auditory cues likely to disambiguate the utterances, hence increasing the natural ambiguity between the two possible percepts, e.g. “l’attache” or “la tache”, the speaker was instructed to produce the material in a neutral way, without trying to induce a preference for one or the other segmentation. A previous study using exactly the same acoustic material (Strauß et al., 2015) demonstrated that it was indeed neutral enough for a word segmentation task because obtained percentages of judgement probability that the participant identified the sound as C-onset word (e.g. “la tache”) were slightly above chance rate.

Among these seventeen stimuli, 5 involved /a/ as the first vowel in the target noun, with a pure vertical tongue-jaw opening gesture, while the first vowel for the 12 other pairs of nouns was a vowel within /ɛ ɛ̃ i ə

ɑ̃ ɔ œ/ involving mainly a horizontal front-back gesture of the tongue and a horizontal round-spread gesture of the lips (as indicated in the last column in Table 1).

In the test, the auditory stimuli were presented through headphones (AKG K242), one at a trial. The stimulus sound intensity was adjusted to a fixed comfortable level for each participant.

****

Table 1: Corpus of French target phrases

No.	V-onset word	C-onset word	Pronunciation	Articulatory direction for the first vowel in the C-onset word (in bold face).
1	l'alarme	la larme	/lal**a**ʀm/	Vertical
2	l'avarice	la varice	/lav**a**ʀis/	
3	l'attraction	la traction	/latʀ**a**ksjɔ̃/	
4	l'amarre	la mare	/lam**a**ʀ/	
5	l'attache	la tache	/lat**a**ʃ/	
6	l'aversion	la version	/lav**ɛ**ʀsjɔ̃/	Horizontal
7	l'atteinte	la teinte	/latɛ̃t/	
8	l'amie	la mie	/lam**i**/	
9	l'affiche	la fiche	/laf**i**ʃ/	
10	l'haleine	la laine	/lal**ɛ**n/	
11	l'attention	la tension	/lat**ɑ̃**sjɔ̃/	
12	l'avenue	la venue	/lav(**ə**)ny/	
13	l'attente	la tente	/lat**ɑ̃**t/	
14	l'amante	la mante	/lam**ɑ̃**t/	
15	l'annotation	la notation	/lan**ɔ**tasjɔ̃/	
16	l'allocation	la location	/lal**ɔ**kasjɔ̃/	
17	l'apesanteur	la pesanteur	/lap**œ**zɑ̃tœʀ/	

****

2.3 Somatosensory stimulation

Somatosensory stimulation was applied by using a small robotic device (PHANToM Premium 1.0, SenSable Technologies). Small plastic tabs (2 × 3 cm in each tab), connected to the robot through thin string, were attached to the skin at a given location (see later), using double sided tape. A given stimulation consisted in a

sinusoidal pulse provided by a half-wave sinusoidal movement at 6 Hz (167 ms duration). This duration was selected as compatible with a typical vowel duration in the acoustic corpus. It was expected that this pattern of somatosensory stimulation would evoke the somatosensory input associated with the production of a given vowel in the sequence. The pulse was applied at a given position in time in the sequence, with eight possible onset timings (P1-P8). Figure 1 represents the temporal relationships between the somatosensory and auditory stimulations. As a reference for timing in all phrases, we first set the onset of stimulation P5 at the envelope peak of the first vowel in the article+noun phrase (e.g. first /a/ in “attache”). Envelope was computed by the root mean square of the amplitude of the acoustic signal. Onsets of the other stimulations were set by 100 ms intervals between two consecutive positions (see Figure 1). Note that no audible sound was produced in the force generation by the robot.

****

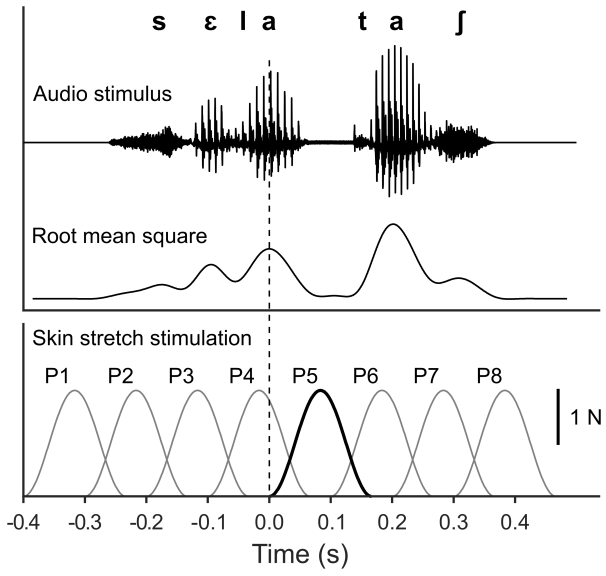


Figure 1: Temporal relationship between the audio stimulus and skin stretch stimulation. (Top) Acoustic signal and its root mean square (RMS) envelope for one example in the acoustic corpus, “C’est l’attache” or “C’est la tache” (/sɛlataʃ/). (Bottom) Temporal pattern of the skin stretch stimulation with eight possible stimulus onset timings (P1-P8). Stimulation “P5”, which was used as a basis to set the onsets of all stimulations, is in thick line, and skin stretch stimulations at other timings are in thin grey line. The vertical dotted line is synchronous with the RMS acoustic peak of the first vowel in the V-onset word.

****

201

202 **2.4 Experimental procedure**

203 Participants were presented with various combinations of acoustic and somatosensory stimuli, detailed here
204 under. Their task was to decide whether the corresponding acoustic stimulus corresponded to either a V-
205 onset word (e.g. “l’attache”) or a C-onset word (e.g. “la tache”), by pressing on the corresponding key on a
206 keyboard as quickly as possible. The experiment comprised 17 acoustical sentences, each one associated to 9
207 somatosensory conditions: 8 conditions with a somatosensory stimulation at one of the 8 possible timings
208 (P1-P8) and one pure audio condition with no somatosensory stimulation (P0). All possible pairs of auditory-
209 somatosensory stimulations (153 pairs = 17 acoustic phrases × 9 somatosensory conditions) were presented
210 in one block in a pseudo random order, with two restrictions. Firstly, the acoustic phrase systematically
211 differed from one trial to the next. Secondly, every nine trials, the whole set of nine somatosensory
212 conditions was tested. The block was presented four times with a short break between blocks. In total, 612
213 trials were recorded per participant.

214

215 Participants were randomly assigned between two groups (20 participants per group) corresponding to two
216 sites for somatosensory stimulation. For the first group, stimulation was applied on both sides of the
217 participant’s mouth, in the upward direction (Face condition). This condition, which is the major focus of the
218 present study, corresponds exactly to what was used in our previous studies (Ito & Gomi, 2007; Ito et al.,
219 2009). For the second group, stimulation was applied on the skin on the left forearm, horizontally towards
220 the hand (Forearm condition). This provided the control site for the experiment. This site was selected for its
221 property that the skin on this part of the body has a sensitivity similar to the orofacial skin in a force
222 judgement task (Ito & Ostry, 2012). The left forearm was selected rather than the right forearm to minimize
223 the possible interactions between hand and mouth representations, known to be strong in the left frontal
224 cortex (see recent reviews in Aboitiz, 2018; Króliczak, Gonzalez, & Carey, 2019) while keeping a good
225 capacity to display temporal information, known to be optimal in the syllable and word range in the right
226 auditory cortex (Giraud & Poeppel, 2012; Poeppel, 2003).

227

228 **2.5 Hypotheses and data analysis**

The working hypothesis of this study is that a pure audio presentation (condition P0) would provide rather ambiguous decisions (around 50% of each response in a pair). On that basis, the first tested assumption is that in the facial skin condition, the temporal position of the stimulation would modulate the perceptual decision, towards a preference for the V-onset word (e.g. “l’attache”) for a stimulation around the first vowel and rather in the other direction (towards the C-onset word, e.g. “la tache”) for later somatosensory stimulations. The second tested hypothesis was that the effect would be weaker or absent in the Forearm condition. If audio-somatosensory integration were based on inferences related to speech production, only a facial skin stretch condition could be relevant, with no effect at all on the forearm. As an alternative, it could also be assumed that audio-somatosensory interactions in this paradigm would be based on pure temporal information independent on the sensory channel providing this information. In this case, the Forearm condition could also lead to variations of the perceptual decision with the temporal position of the somatosensory stimulation, though possibly with less efficiency than the Face condition.

To assess these assumptions, the judgement probability that the participant identified the presented word as C-onset word (e.g. “la tache”) was calculated for each phrase, each stimulation condition (with 4 repetitions per case) for each participant in each of the two groups. Then, a first global statistical analysis was carried out by applying a Linear Mixed-Effects Model with the R software (version 3.5.3) (R Core Team, 2019), with a fixed between-subject factor stimulation *Site* (Face vs. Forearm) and a fixed within-subject factor somatosensory stimulation *Onset* (P0-P8), with participants as a random factor. We used the lme function from the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019) for global analysis with the following formula: Probability ~ *Site* * *Onset*, random=~1|Subject. In this analysis, we specifically focused on an interaction (*Site* × *Onset*) effect which was tested by comparing between models with and without the interaction term. Post-hoc tests were carried out using multiple comparisons with Bonferroni correction. We used the glht function from the multcomp package (Hothorn, Bretz, & Westfall, 2008) to compare all possible combinations separately in each *Site*.

In a follow-up analysis, we further examined how the articulatory characteristics involved in the specific configuration associated with each phrase affected the effect of the somatosensory stimulation on word

segmentation. Based on our findings that the somatosensory effect was limited to the Face condition (see Section 3.1), we applied this analysis to this condition alone. In the test stimuli, the first vowel was always "a", but the following vowel was varied (see Table 1). We divided the stimulus words into two groups according to articulatory characteristics in the following vowels. In the first group ("Vertical"), we considered the words in which the second vowel was a low open /a/ involving only a vertical opening movement compatible with the direction of the somatosensory stimulation. In the second group ("Horizontal"), we considered all the other words which actually involve two possible horizontal movements respectively outwards in spreading gestures (e.g. /i/ in "affiche" or /ɛ/ in "aversion") or inwards in rounding (e.g. /œ/ in "l'apesanteur" or /ɔ/ in "l'allocation"). We assumed that there should be lesser variations of lexical decision with somatosensory stimulation timing in the second group, since the direction of the stimulation is not compatible with the direction of the corresponding speech articulation for producing the second vowel. To test this last assumption, a Linear Mixed-Effects Model was applied in the Face condition with fixed within-subject factors *Group* (Vertical vs. Horizontal) and stimulation *Onset* (P0-P8), with participants as a random factor. We applied the following formula: Probability ~ *Group* * *Onset*, random=~1|Subject.

3. Results

3.1 Face vs. Forearm

Figure 2 presents the average judgement probability in the word identification task across somatosensory conditions. In the Face condition, the judgement probability appears to vary with the timing of the somatosensory stimulation (Figure 2A). When the somatosensory stimulation leads the first vowel (=P3), the judgement probability reaches the smallest value overall. When the somatosensory stimulation is close to the second vowel (=P6), the judgement probability is larger. In the Forearm condition, the average judgement probability does not appear to depend much on timing, staying around chance (50%) for all timing conditions of the somatosensory stimulation (Figure 2B).

****

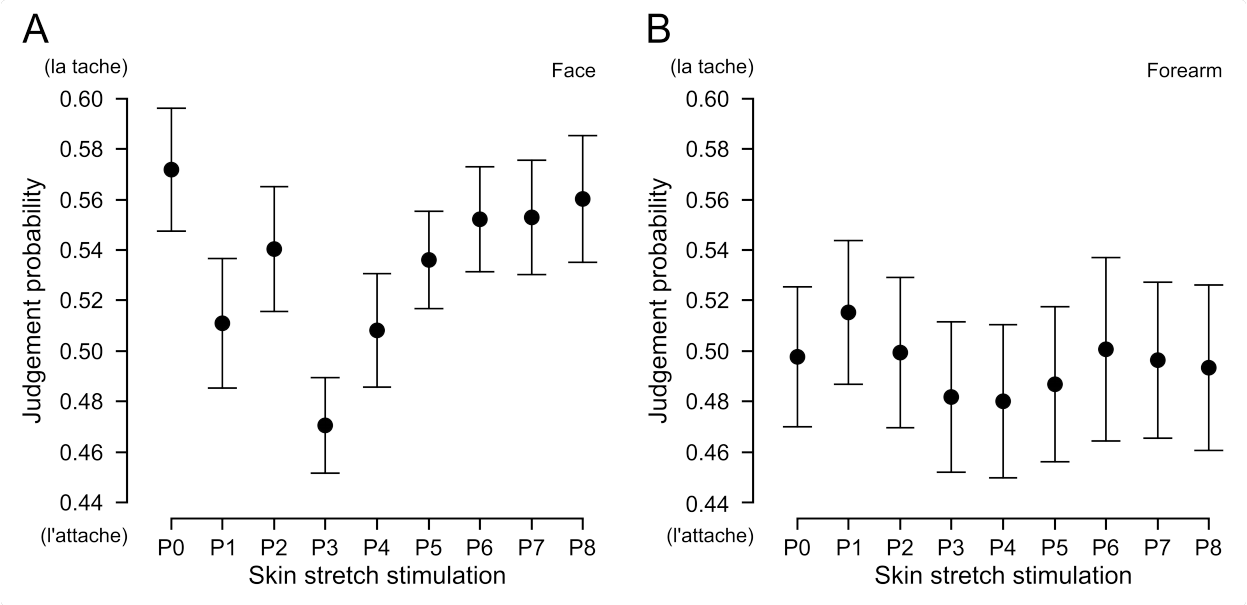


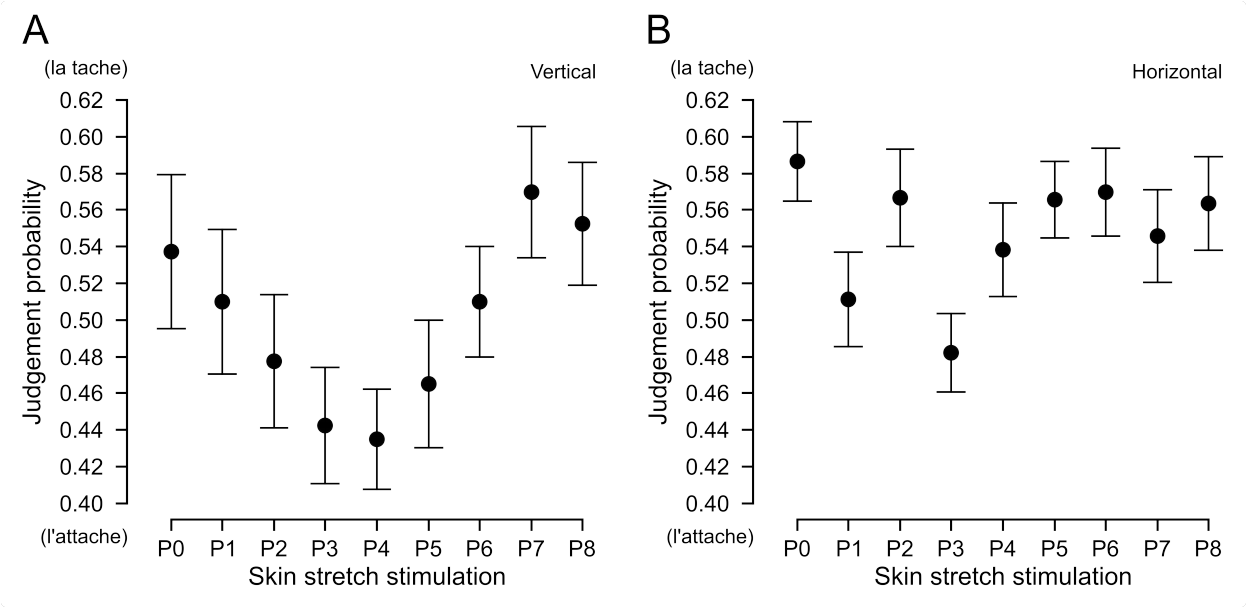
Figure 2: Average judgement probability in the Face (A) and the Forearm condition (B). The horizontal axis represents the type of skin stretch stimulation (P0-P8). The vertical axis represents the judgement probability that participants identified the audio stimulus as a C-onset word (e.g. “la tache”), averaged over participants in the corresponding group. Error bars represent standard errors across participants.

****

The Linear Mixed-Effects Model analysis provides no significant effect of stimulation *Site* (χ^2 (1) = 1.42, p > 0.2329), but a significant effect of stimulation *Onset* (χ^2 (8) = 26.00, p < 0.0011) and, most importantly, a

295 significant interaction between *Site* and *Onset* (χ^2 (8) = 18.04, $p < 0.0210$), indicating that the timing of the
296 somatosensory stimulation affected differently the participants' responses depending on the stimulation site.
297 Post-hoc tests confirm that there is a significant effect of the somatosensory stimulation onset in the Face
298 condition, with significant differences between P0 and P3 ($p < 0.001$), P3 and P6 ($p < 0.003$), P3 and P7 ($p <$
299 0.002), P3 and P8 ($p < 0.001$) and P2 and P3 ($p < 0.026$). On the other hand, there is no significant difference
300 between any onset value in the Forearm condition (all p values > 0.9). This suggests that somatosensory
301 inputs affected lexical decision in the Face condition, and particularly when the skin stretch stimulation was
302 applied around P3, but not in the Forearm condition. Detailed tables for this statistical analysis are provided
303 as supplementary data.
304
305 **3.2 Vertical vs. Horizontal word groups**
306 Figure 3 presents the variations of lexical judgement with somatosensory stimulation onset in the Face
307 condition, separately for the Vertical (Figure 3A) and the Horizontal (Figure 3B) word groups (see Table 1).
308 The pattern of responses appears more regular with larger variations with stimulation onset in the Vertical
309 group, with a gradual decrease of the amount of judgement probability around P3-P4, followed by a gradual
310 increase up to a value larger than for P0, around P7. The values for the Horizontal group are more irregular,
311 with a strong decrease of judgement probability for P3 and to a lesser degree P1, but basically no variation
312 for other stimulation onset values.
313
314
315
316
317
318
319
320
321
322

323 **********



324

325 **Figure 3: Average judgement probability for the two articulatory direction groups: "Vertical" (A) and**

326 **"Horizontal" (B) in the Face condition.** Same presentation as in Fig. 2.

327 **********

328

329 The Linear Mixed-Effects Model analysis provides a significant main effect of stimulation *Onset* (χ^2 (8) =

330 32.56, $p < 0.0001$) as already mentioned in the global analysis, together with a significant effect of *Group* (χ^2

331 (1) = 18.12, $p < 0.0001$). The interaction between *Onset* and *Group* is close to significant (χ^2 (8) = 15.26, $p =$

332 0.0542). Altogether, hence, the pattern of responses to the somatosensory stimulation differs between groups,

333 with probably a larger range of variations of the lexical response with stimulation onset in the Vertical word

334 group. Detailed tables for this statistical analysis are also available as supplementary data.

335

336

337

4. Discussion

The results in Section 3 provide clear evidence relative to the hypotheses introduced in Sections 1.2 and 2.5. Firstly, there was indeed an effect of the timing of the somatosensory stimulation on lexical decision when the stimulation was applied on the face, though not on the forearm. Furthermore, the effect appeared to be significantly different depending on the articulatory nature of the vowel in the tested word (Vertical vs. Horizontal). We will first propose a global interpretation of these data related to the role of articulatory information in perception. Then we will discuss what could be the consequences of these experimental data for a global theory of speech perception. We will finally raise some limitations and perspectives for further studies.

4.1 Modulation of auditory word segmentation by articulatory-compatible somatosensory stimulation

The main finding of the present study is that somatosensory inputs associated with facial skin deformation modulate the perception of lexical information in French. The effect was mainly induced in relation with the timing of the somatosensory stimulation relative to vowel targets in the acoustic input. Globally, the pattern of somatosensory effects appears closely related to the nature of the underlying articulatory movements associated with the production of the corresponding acoustic sequences. This was shown by three important aspects of the somatosensory modulation of auditory word segmentation as follows.

4.1.1 An audio-somatosensory timing compatible with the dynamics of speech gestures

The data in Figure 2 and Section 3.1 show that the timing of the facial somatosensory stimulation relative to the stimulus sound does play a role in the segmentation process for lexical perception. The word category can be taken as a proxy of the segmentation process. Let us assume that the decision for V-onset words (e.g. “l’attache” /l#ataʃ/) entails that a major prosodic accent has been perceived by the participants at the position of the first vowel /a/, while the decision for C-onset word (e.g. “la tache” /la#taʃ/) entails that a major prosodic accent has been rather perceived at the position of the second vowel. Then the underlying interpretation of the data in the Face condition is that the vertical movement of the face induced by the facial skin stretch stimulation is interpreted by the subjects as a clue that the following vowel is indeed accentuated, hence the corresponding decision about word segmentation. Indeed, when the stimulation was

366 applied before the first vowel /a/ around P3, the perception was biased towards V-onset words, while if it
367 was applied after the first /a/ but before the second vowel, the perception was rather biased towards C-onset
368 words.
369
370 The fact that the largest modulation was obtained when the somatosensory stimulation onset led the
371 corresponding acoustic peak (P3) is consistent with anticipatory mechanisms in speech production. The onset
372 of a vocalic gesture can start at least 100 ms and up to 400 ms before the vowel climax (e.g. Noiray et al.,
373 2011). This anticipatory gesture may be seen before it is heard, hence the asymmetry of the temporal window
374 of audiovisual integration (van Wassenhove, Grant, & Poeppel, 2007) and the observation that the maximal
375 effect of the incongruent visual input in the McGurk effect (MCGURK & MACDONALD, 1976) may occur
376 for a visual advance on the sound (Munhall, Gribble, Sacco, & Ward, 1996). The peak of perceptual
377 modulation by a somatosensory stimulus at P3 is also in line with the observation by Ito, Gracco, & Ostry
378 (2014) that the change of cortical potentials by auditory-somatosensory interaction is induced specifically
379 when somatosensory inputs precede auditory inputs.
380
381 **4.1.2 A somatosensory effect specific of facial stimulation and absent for forearm stimulation**
382 A second crucial finding in Figure 2 and Section 3.1 is that a somatosensory stimulation applied on the
383 forearm produced no effect on the auditory word segmentation process. This is unlikely to be due to a lack of
384 sensitivity of this region of the body, since Ito & Ostry (2012) showed similar sensitivity of the face and the
385 forearm in a force judgement task. Nor is it related to a lack of ability of the right cortex hemisphere, dealing
386 with left forearm afferences, to process speech temporal information, considering the “asymmetric sampling
387 in time” hypothesis by Poeppel (2003). Indeed, this hypothesis, later confirmed in a number of studies (see a
388 review in Giraud & Poeppel (2012), claims that the right auditory cortex would preferentially extract
389 information from long integration windows, typically 150-to-250 ms wide, corresponding exactly to the
390 temporal range involved in the present study, while the left auditory cortex would rather be in charge of
391 shorter temporal windows around 25 to 40 ms.
392

The difference between results obtained for the forearm vs. facial stimulation is rather due to a stimulus-response compatibility effect, in which the forearm would not be relevant for the task at hand. This kind of compatibility effects has been observed by Gick & Derrick (2009) who showed that air puff stimulation on the hand or neck while hearing aspirated (/pa/ or /ta/) or unaspirated sounds (/ba/ or /da/) modulated perceptual judgements, while no effect was induced by simple tapping. More direct evidence for the specificity of the relation between face and speech is provided by Ito & Ostry (2012) who showed that facial skin sensation was altered by listening to speech sounds, but not by listening to non-speech sounds. Their study also showed that skin sensations of forearm and palm were not altered by listening to speech sounds, contrary to face sensations. Similarly, somatosensory event-related potentials induced by facial skin deformation were modulated by presenting speech sounds, but not by non-speech sounds or noise (Ito, Johns, & Ostry, 2013), whereas simple lip tapping stimulation during speech perceptual processing did not affect magnetoencephalographic response changes (Möttönen, Järveläinen, Sams, & Hari, 2005).

The lack of forearm effect provides evidence against the "soft" version of our second hypothesis in Section 2.5, that there could exist an effect of timing information applied to the forearm, independently on the speech-face compatibility. While there is timing information available on the forearm in the corresponding condition, the participants happen to neglect this pure temporal information for word segmentation, confirming that the effect of facial skin stimulation is indeed due to an articulatory interpretation of the stimulation, impossible with the stimulation on the left forearm.

4.1.3 Facial somatosensory stimulation seems more compatible with vertical than with horizontal articulatory gestures

Finally, there was a weak trend that there would be more effect of the stimulation for words involving a vertical (opening) rather than a horizontal (rounding/spreading) movement for the vowel after the first /a/, and the pattern of effects was clearly different between the two groups (Figure 3). This can also be compatible with the importance of the congruence between the orofacial stimulation and the associated orofacial speech gesture. Facial skin deformation in speech production may occur in various directions depending on the uttered word (Vatikiotis-Bateson, Kuratate, Kamachi, & Yehia, 1999). Ito et al. (2009)

showed that horizontal displacements applied on the skin lateral to the mouth could not modify the perception of opening gestures towards /a/ or /ɛ/. Importantly, upward displacements of the face are compatible with such opening gestures (Ito & Ostry, 2012), hence direction of stimulation seems crucial here, rather than the precise sense of the stimulation along the corresponding direction – though Ogane, Schwartz, & Ito (2019) showed that the precise amplitude of the orofacial somatosensory stimulation is of weak importance in the modulation of lexical perception. In the present case, while the vertical stimulation around P3 is compatible with the opening gesture for the first /a/ independently on the following word, it is compatible with the second vowel only if it is an /a/, in the Vertical group. This could well explain the lack of effect of the stimulation after P3 in the Horizontal group, while the pattern of modulation seems to extend until P7-P8 in the Vertical group (Figure 3).

Therefore, the pattern of modulation of lexical decision observed in the present data seems to support the hypothesis that French listeners can differentiate the phrases with elision by information about the respective strength of articulation of the first and second vowels, provided by the somatosensory inputs associated with facial skin deformation. Such kinesthetic information about speech production can be provided by orofacial cutaneous mechanoreceptors (Ito & Ostry, 2010; Johansson et al., 1988). Because of the deformation of the lower face area during opening speech movements, cutaneous mechanoreceptors in the skin around the mouth might be predominant in the detection of articulatory movement. The current somatosensory stimulation likely induced somatosensory inputs related to the listener's expectations about her/his own speaking gestures, in line with theories invoking the role of information related to speech production in the speech perception process. This will be the topic of the next section.

4.2 Consequences for a theory of speech perception

This study adds to a number of data showing that the phonetic interpretation of multisensory stimulation is related to the underlying articulatory knowledge available to the tested participant. More specifically, it extends previous studies on the role of the somatosensory system in speech perception, e.g. Ito et al. (2009), Gick & Derrick (2009), to a novel paradigm that is segmentation for lexical access.

In spite of the increasing agreement that articulatory processes may intervene in speech perception (see Section 1.1), there remains a large range of different views on the nature of this intervention. The historical pioneer view from the Motor Theory of Speech Perception (Liberman et al., 1967; Liberman & Mattingly, 1985) was that speech perception was based on underlying motor representations and motor gestures. Perceptuo-motor theories modified this view by rather considering a mix of auditory and motor processes in speech perception (Schwartz et al., 2012; Skipper et al., 2007). Skipper et al. (2007) posited an analysis-by-synthesis process targeting motor programs associated with the phoneme. Stokes et al. (2019) claimed that the motor system has at best a modulatory role, probably at the level of phonemic processing.

The present data probably set the balance towards a larger and more integrative role of the motor system in speech perception. Indeed, it appears that articulatory knowledge elicited by facial somatosensory stimulation intervenes at the level of the global processing of the acoustic stream, by orienting segmentation towards some parts rather than others. A similar claim has been done by Basirat, Schwartz, & Sato (2012) in their study of the verbal transformation effect arguing that the motor system was likely to intervene at the stage of chunking the acoustic stream on the basis of articulatory underlying trajectories. It is also consistent with Remez et al. (1994) that the articulatory/motor nature of the speech stream contributes to gluing the various pieces of information together in the perceptual analysis and decoding process. Strauß & Schwartz (2017) have further proposed that the syllabic rhythm could emerge as an audio-motor construction. These views are integrated in the framework of the Perception-for-Action-Control Theory (Schwartz et al., 2012), which proposes that the whole speech analysis process should be conceived as a perceptuo-motor process in which auditory shaping and motor knowledge would be constantly combined in the analysis and decoding process connecting the sensory inputs with the lexical knowledge in the human brain.

4.3 Limitations and perspectives

The present evidence that facial skin deformation associated with speech articulatory movements can intervene in segmentation before lexical access is based on a very specific process available in French, provided by elision between determinants and nouns. Hence it may be wondered whether similar evidence could be found in other languages which do not involve the same kind of phonological process. Our current

assumption is that similar effects should be obtained in other languages as long as different articulatory movements provide a clue for the disambiguation in lexical decisions. Still this remains to be demonstrated, and a first stage would consist in finding similar configurations in other languages. Of course, somatosensory interferences in segmentation are not expected in situations where no clear correlate associated to articulation is available in production, and likely to be exploited reciprocally in perception.

The present study exploits a rather limited set of facial somatosensory stimulation, with a single direction of movement (vertical upwards) and a fixed intensity of stimulation. This is another clear limitation of this study, considering that vowels differ both in the directionality and the amplitude of their trajectory from the previous consonant. Evidence that the amplitude of the somatosensory effect changed depending on the first vowel in the C-onset word (Section 3.2) confirms the need for some amount of matching with the stimulation. In a recent study (Ogane, Schwartz, & Ito, 2019) we explored the effect of two sequential skin stretches with a different amplitude in contrast with the single pulse stimulation used in the present study. We did not at this point obtain a significant effect of stimulations with contrasted amplitudes, probably because of the difficulty of selecting adequate contrasted somatosensory stimulations able to represent efficiently a contrast between two vowels. Since the actual articulatory movements are far more complex than simple sequences of vertical gestures, more realistic patterns of somatosensory stimulation (e.g. multiple stimulation pulses with different directions of stimulation and varying amplitudes) are required for further more detailed and precise investigation of possible perceptual modulations by somatosensory inputs.

5. Conclusion

This study showed that the lexical perception of a given sequence in French, involving ambiguous word segmentation, can be significantly modified by applying a somatosensory input on the facial skin. The judgement was systematically biased towards one or the other segmentation depending on the timing of the somatosensory input. Importantly, this effect was specifically induced by a stimulation on the facial skin, but not on the skin elsewhere than the face (forearm). In a follow-up analysis, we also found that the effect of the somatosensory stimulation in word segmentation was different and globally larger and more coherent in phrases involving a vertical articulatory movement for the first vowel in the C-onset word than for vowels involving horizontal movements of the tongue or lips. Altogether, these data are consistent with the proposal that somatosensory information arising from the facial skin is involved not only in phonetic perception, but also at higher stages in speech perception, that is, at the level of segmentation of the acoustic stream for lexical access. This provides an important argument to the conception of a close connection between perceptual and motor processes in the whole speech processing chain in speech communication.

Acknowledgements

We thank Nathan Mary and Dorian Deliquet for data collection and analysis and Silvain Gerber for statistical analysis.

Funding information

This work was supported by the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 339152). This work was supported by grant ANR-15-IDEX-02 CDP NeuroCoG.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

R.O., J.-L.S. and T.I. designed research. R.O. and T.I. performed the experiment. R.O., J.-L.S. and T.I. analyzed data. R.O., J.-L.S. and T.I. wrote the paper. R.O., J.-L.S. and T.I. confirmed the final version.

References

Aboitiz, F. (2018). A Brain for Speech. Evolutionary Continuity in Primate and Human Auditory-Vocal Processing. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00174

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903–1909. https://doi.org/10.1177/0956797610389192

Basirat, A., Schwartz, J., & Sato, M. (2012). Perceptuo-motor interactions in the perceptual organization of speech: evidence from the verbal transformation effect. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 965–976. https://doi.org/10.1098/rstb.2011.0374

Borrie, S. A., & Schäfer, M. C. M. (2015). The Role of Somatosensory Information in Speech Perception: Imitation Improves Recognition of Disordered Speech. *Journal of Speech Language and Hearing Research*, *58*, 1708–1716. https://doi.org/10.1044/2015

Connor, N. P., & Abbs, J. H. (1998). Movement-related skin strain associated with goal-oriented lip actions. *Experimental Brain Research*, *123*(3), 235–241. https://doi.org/10.1007/s002210050565

D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I., & Craighero, L. (2011). Tongue corticospinal modulation during attended verbal stimuli: Priming and coarticulation effects. *Neuropsychologia*, *49*(13), 3670–3676. https://doi.org/10.1016/j.neuropsychologia.2011.09.022

Dohen, M., Lœvenbruck, H., Cathiard, M.-A., & Schwartz, J.-L. (2004). Visual perception of contrastive focus in reiterant French speech. *Speech Communication*, *44*(1–4), 155–172. https://doi.org/10.1016/j.specom.2004.10.009

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research*, *12*(2), 423–425. https://doi.org/10.1044/jshr.1202.423

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*, 399–402. https://doi.org/10.1046/j.0953-816x.2001.01874.x

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*(8), 1207–1223. https://doi.org/10.1080/01690965.2012.701758

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French.

559 *Journal of Phonetics*, *29*(2), 109–135. https://doi.org/10.1006/jpho.2000.0114
560 Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*, 502–504.
561 https://doi.org/10.1038/nature08572
562 Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational
563 principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063
564 Grabski, K., Schwartz, J.-L., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., … Sato, M. (2013). Shared and
565 distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics*, *26*(3), 384–
566 408. https://doi.org/10.1016/j.jneuroling.2012.11.003
567 Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of
568 spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197.
569 https://doi.org/10.1121/1.1288668
570 Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models.
571 *Biometrical Journal*, *50*(3), 346–363. https://doi.org/10.1002/bimj.200810425
572 Ito, T., & Gomi, H. (2007). Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in
573 speech. *NeuroReport*, *18*(9), 907–910. https://doi.org/10.1097/WNR.0b013e32810f2dfb
574 Ito, T., Gracco, V. L., & Ostry, D. J. (2014). Temporal factors affecting somatosensory-auditory interactions
575 in speech processing. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.01198
576 Ito, T., Johns, A. R., & Ostry, D. J. (2013). Left Lateralized Enhancement of Orofacial Somatosensory
577 Processing Due to Speech Sounds. *Journal of Speech Language and Hearing Research*, *56*(6), 1875–
578 1881. https://doi.org/10.1044/1092-4388(2013/12-0226)
579 Ito, T., & Ostry, D. J. (2010). Somatosensory Contribution to Motor Learning Due to Facial Skin
580 Deformation. *Journal of Neurophysiology*, *104(3)*, 1230–1238. https://doi.org/10.1152/jn.00199.2010
581 Ito, T., & Ostry, D. J. (2012). Speech sounds alter facial skin sensation. *Journal of Neurophysiology*, *107*(1),
582 442–447. https://doi.org/10.1152/jn.00029.2011
583 Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the*
584 *National Academy of Sciences*, *106*(4), 1245–1248. https://doi.org/10.1073/pnas.0810063106
585 Johansson, R. S., Trulsson, M., Olsson, K. Å., & Abbs, J. H. (1988). Mechanoreceptive afferent activity in
586 the infraorbital nerve in man during speech and chewing movements. *Experimental Brain Research*, *72*,

209–214. https://doi.org/10.1007/BF00248519

Kim, J., & Davis, C. (2004). Investigating the audio-visual speech detection advantage. *Speech Communication*, *44*, 19–30. https://doi.org/10.1016/j.specom.2004.09.008

Króliczak, G., Gonzalez, C. L. R., & Carey, D. P. (2019). Editorial: Manual Skills, Handedness, and the Organization of Language in the Brain. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00930

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the Human Speech Motor System Drives Changes in Speech Perception. *Journal of Neuroscience*, *34*(31), 10339–10346. https://doi.org/10.1523/JNEUROSCI.0108-14.2014

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*. https://doi.org/10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mitchel, A. D., & Weiss, D. J. (2014). Visual speech segmentation: using facial cues to locate word boundaries in continuous speech. *Language, Cognition and Neuroscience*, *29*(7), 771–780. https://doi.org/10.1080/01690965.2013.791703

Möttönen, R., Järveläinen, J., Sams, M., & Hari, R. (2005). Viewing speech modulates activity in the left SI mouth cortex. *NeuroImage*, *24*(3), 731–737. https://doi.org/10.1016/j.neuroimage.2004.10.011

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, *58*(3), 351–362. https://doi.org/10.3758/BF03206811

Noiray, A., Cathiard, M.-A., Ménard, L., & Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America*, *129*(1), 340–349. https://doi.org/10.1121/1.3518452

Nordin, M., & Hagbarth, K.-E. (1989). Mechanoreceptive units in the human infra-orbital nerve. *Acta Physiologica Scandinavica*, *135*(2), 149–161. https://doi.org/10.1111/j.1748-1716.1989.tb08562.x

Ogane, R., Schwartz, J.-L., & Ito, T. (2019). Orofacial Somatosensory Effects for the Word Segmentation

615 Judgement. In *Proceedings of International Congress of Phonetic Sciences (ICPhS) 2019*. Melbourne,
616 Australia.
617 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). nlme: Linear and Nonlinear Mixed
618 Effects Models. Retrieved from https://cran.r-project.org/package=nlme
619 Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization
620 as 'asymmetric sampling in time.' *Speech Communication*, *41*(1), 245–255.
621 https://doi.org/10.1016/S0167-6393(02)00107-3
622 Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor
623 cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*,
624 *103*(20), 7865–7870. https://doi.org/10.1073/pnas.0509989103
625 R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R
626 Foundation for Statistical Computing. Retrieved from https://www.r-project.org/
627 Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the Perceptual Organization
628 of Speech. *Psychological Review*, *101*(1), 129–156. https://doi.org/10.1037/0033-295X.101.1.129
629 Sato, M., Grabski, K., Glenberg, A. M., Brisebois, A., Basirat, A., Ménard, L., & Cattaneo, L. (2011).
630 Articulatory bias in speech categorization: Evidence from use-induced motor plasticity. *Cortex*, *47*,
631 1001–1003. https://doi.org/10.1016/j.cortex.2011.03.009
632 Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory
633 (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336–354.
634 https://doi.org/10.1016/j.jneuroling.2009.12.004
635 Sell, A. J., & Kaschak, M. P. (2009). Does visual speech information affect word segmentation? *Memory*
636 *and Cognition*, *37*(6), 889–894. https://doi.org/10.3758/MC.37.6.889
637 Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds
638 following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113.
639 https://doi.org/10.1121/1.3058638
640 Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices:
641 How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception. *Cerebral*
642 *Cortex*, *17*(10), 2387–2399. https://doi.org/10.1093/cercor/bhl147

643 Spinelli, E., Grimault, N., Meunier, F., & Welby, P. (2010). An intonational cue to word segmentation in
644 phonemically identical sequences. *Attention, Perception, & Psychophysics*, *72*(3), 775–787.
645 https://doi.org/10.3758/APP.72.3.775
646 Spinelli, E., Welby, P., & Schaegis, A.-L. (2007). Fine-grained access to targets and competitors in
647 phonemically identical spoken sequences: the case of French elision. *Language and Cognitive*
648 *Processes*, *22*(6), 828–859. https://doi.org/10.1080/01690960601076472
649 Stål, P., Eriksson, P.-O., Eriksson, A., & Thornell, L.-E. (1990). Enzyme-histochemical and morphological
650 characteristics of muscle fibre types in the human buccinator and orbicularis oris. *Archives of Oral*
651 *Biology*, *35*(6), 449–458. https://doi.org/10.1016/0003-9969(90)90208-R
652 Stokes, R. C., Venezia, J. H., & Hickok, G. (2019). The motor system's [modest] contribution to speech
653 perception. *Psychonomic Bulletin & Review*, *26*(4), 1354–1366. https://doi.org/10.3758/s13423-019-
654 01580-2
655 Strauß, A., Savariaux, C., Kandel, S., & Schwartz, J.-L. (2015). Visual lip information supports auditory
656 word segmentation. In *FAAVSP 2015*. Vienna, Austria.
657 Strauß, A., & Schwartz, J.-L. (2017). The syllable in the light of motor skills and neural oscillations.
658 *Language, Cognition and Neuroscience*, *32*(5), 562–569.
659 https://doi.org/10.1080/23273798.2016.1253852
660 Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of*
661 *the Acoustical Society of America*, *26*, 212–215. https://doi.org/10.1121/1.1907309
662 Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral
663 premotor cortex to speech perception. *NeuroImage*, *57*(4), 1561–1571.
664 https://doi.org/10.1016/j.neuroimage.2011.05.067
665 van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-
666 visual speech perception. *Neuropsychologia*, *45*, 598–607.
667 https://doi.org/10.1016/j.neuropsychologia.2006.01.001
668 Vatikiotis-Bateson, E., Kuratate, T., Kamachi, M., & Yehia, H. (1999). Facial deformation parameters for
669 audiovisual synthesis. In *AVSP'99* (pp. 118–122). Santa Cruz, CA, USA.
670 Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system

involved in speech production. *Neuropsychologia*, *41(8)*, 989–994. https://doi.org/10.1016/S0028-3932(02)00316-0

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701–702. https://doi.org/10.1038/nn1263