OROFACIAL SOMATOSENSORY EFFECTS FOR THE WORD SEGMENTATION JUDGEMENT

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

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France
[2] Haskins Laboratories, CT, USA
* Institute of Engineering Univ. Grenoble Alpes
{rintaro.ogane, jean-luc.schwartz, takayuki.ito}@gipsa-lab.grenoble-inp.fr

ABSTRACT

Word segmentation is one of the initial processes for lexical perception. While visual inputs can help it in acoustically ambiguous situations, the effect of orofacial somatosensory inputs to this process is unknown. We here tested how orofacial somatosensory inputs affect word segmentation for lexical perception. We carried out identification tests using a French phrase consisting of a definitive and a noun, segmented differently according to the place of the accents in the phrase. In the test applying somatosensory stimulation at various timings along the phrase with neutral accent, we found that the lexical perception was significantly and systematically biased depending on the somatosensory stimulus timing. This bias effect was not seen when two somatosensory stimuli were applied to emphasize one accent position rather than the other by changing force amplitude between two positions. The results show and quantify the role the orofacial somatosensory system plays in lexical perception.

Keywords: lexical perception, temporal timing, multisensory integration, speech production.

1. INTRODUCTION

In speech communication, access to lexical information involves segmentation and decoding processes which both depend on contextual information. Indeed, coarticulatory processes classically modify the acoustic content of a given phonological unit, but may also intervene to blur or enhance the segmentation process, crucial for lexical access. 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 and decoding processes.

In a more general statement, speech perception is an interactive process with multiple sensory modalities and probably crucial perceptuo-motor connections [10]. Recent finding provides evidence that somatosensory inputs associated with orofacial gestures may modify the perception of speech sounds [5]. The speech-like somatosensory inputs were produced by skin stretch perturbation based on the findings that facial cutaneous mechanoreceptors provide articulatory information [4,6]. Although phonetic boundary in vowel perception were systematically modulated depending on the manner of the facial skin deformation, it has never been evaluated whether these effects could go up to the level of lexical access in speech comprehension.

The current project aims to examine whether the processing of lexical information concerning word segmentation can be influenced by somatosensory inputs associated with facial skin stretch. Our assumption is that somatosensory inputs could intervene in the segmentation process and hence modify the lexical decision. To test this assumption, we exploited a specific material in French, that is a phrase consisting of a definitive and a noun, segmented differently according to the place of the accents in the phrase. We carried out two experiment focusing on the timing of one somatosensory input relative to the target auditory phrase (Experiment 1) and on the difference in force amplitude of two somatosensory inputs expected to emphasize one vowel relative to the other in the auditory phrase (Experiment 2).

2. METHODS

2.1. Participants

Thirty-one native French speakers (ten males and twenty-one females, mean age ± *SD*: 26.84 ± 7.75 years old) participated in the experiment (twenty for Experiment 1 and eleven for Experiment 2). They had no record of neurophysiological issues for hearing and for orofacial sensation. The protocols of these experiments were approved by the Grenoble Comité d'Ethique pour les Recherches Non Interventionnelles (CERNI). All participants signed the consent form.

2.2. Auditory identification test

An auditory identification test assessing word recognition in relation with word segmentation was

carried out in both Experiments 1 and 2. As auditory stimuli, we focused on "elision" between a definite article and a noun in French. A specific pair of French nouns have the same pronunciation when they are pronounced with a definite article (e.g. "l'affiche" /l#afiʃ/ ["the poster"] and "la fiche" /la#fiʃ/ ["the card"]; here "#" indicates a word boundary), but can be differentiated by hyper-articulation for the production of the first vowel in each word. Seventeen French target utterances were tested, of the form /laCV.../ or /laCCV.../, preceded by a carrier phrase "C'est" ["This is"].

The auditory stimuli were recorded by one native male speaker of French in three speaking accent conditions. The first recording involved a neutral accent without adding hyper-articulation on any single vowel (S_a0), that is, increasing the ambiguity of the utterance between the two possible interpretations. The two other accent conditions (S_a1 and S_a2) were focused on one or the other interpretation by putting an accent on either the first vowel (e.g. /a/ in l'affiche) or the second vowel (e.g. /i/ in l'affiche). This acoustic focus did modify word segmentation and lexical decision in a previous study on the same material [11]. Experiment 1 exploited only the neutral accent condition (S_a0), while the three speaking accents (S_a0, S_a1, and S_a2) were involved in Experiment 2.

The auditory stimulus was presented through headphones (AKG K242). The sound pressure level was adjusted to a comfortable level for each participant. One trial consisted in the presentation of one specific phrase, for which the participant's task was to identify which word (e.g. "l'affiche" or "la fiche") was presented by pressing a key on a keyboard as quickly as possible.

2.3. Somatosensory stimulations

We applied facial skin deformation by somatosensory stimulation as done in the reference study in the field [5]. A robotic device (PHANToM Premium 1.0, SenSable Technologies) was used to generate a stimulation force in synchronization with the auditory stimulus presentation. Small plastic tabs were attached to both sides of the participant's mouth. A stimulation force was applied in the upward direction. A half-wave 6-Hz sinusoidal pattern was used, providing a 167 ms duration compatible with a typical vowel production in the current acoustic material.

2.4. Experimental procedures

2.4.1. Experiment 1

This experiment aims to examine whether the lexical information processing can be modified by applying a single somatosensory facial skin stimulation at various temporal places relative to the first vowel production in the auditory stimuli.

Figure 1 shows a representative example of a temporal relationship between the audio stimulus (which could be either "C'est l'affiche" or "C'est la fiche") and skin stretch stimulation. We tested 8 different timings of somatosensory stimulation onset (P_t1 to P_t8) in order to cover the entire audio stimulus. We set P_t5 at the time of RMS peak for the first vowel in auditory stimuli (e.g. "a" in the "affiche") (see the vertical dashed line in Figure 1). The other onsets were set with 100 ms intervals based on P_t5. We also tested the condition in the absent of skin stretch stimulation (P_t0).

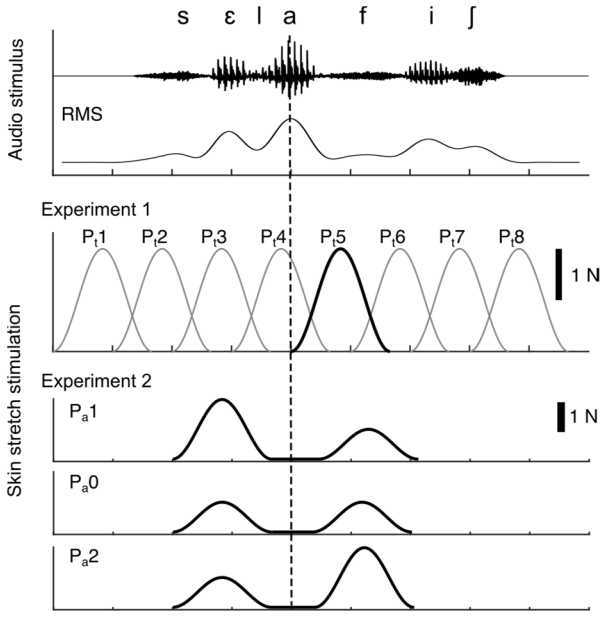
One block consisted of all 9 conditions (P_t0 to P_t8). The order of conditions and auditory stimuli was randomized. A short break was taken every 17 blocks. The total number of trials was 612 (17 French sentences * 9 skin stretch stimulation * 4 repetitions), for a total duration of 30 minutes.

2.4.2. Experiment 2

This experiment aims to examine the potential role of differences in amplitude of somatosensory stimulations in lexical information processing related to word segmentation.

We applied two somatosensory stimulations at timings corresponding respectively to the first and second vowels in the target utterance. The onset of each somatosensory stimulation was set 200 ms

Figure 1: Representative example of a temporal relationship between the audio stimulus, RMS value sequence and skin stretch stimulation. The vertical dashed line represents the time of RMS peak for the first vowel in the stimulus sound.



before the RMS peak of the target vowel, based on the findings of Experiment 1. We controlled the relative amplitude of the two stimulations and tested three patterns of skin stretch stimulation (see the bottom of Figure 1). These patterns exploited two force amplitudes, with a base amplitude set at 1 N and a greater amplitude set at 2 N. In P_a1 condition, the greater stretch amplitude came first and the base one followed. P_a2 condition was in the opposite order with the base first and the greater one following. In P_a0 condition, both stimuli were set at the base amplitude.

One block consisted of 9 experimental conditions [3 speaking accent (S_a0, S_a1 and S_a2) * 3 skin stretch stimulation (P_a0, P_a1 and P_a2)]. The order of conditions and auditory stimuli was randomized. A short break was taken every 17 blocks. The total number of trials was 612 (17 French sentences * 9 experimental conditions * 4 repetitions).

2.5. Data analysis

The probability of response of the type "la + C(C)V", e.g. "la fiche", was calculated for each participant, for each facial skin stretch condition.

For Experiment 1, judgement probability from P_t1 to P_t8 were normalized by dividing by the probability for P_t0. All data were transformed into Z scores. We applied Linear Mixed-Effects Model (LMM) with R (version 3.5.1) [9]. The fixed factor was the Stimulation condition (P_t1 to P_t8) and the random factor was the Participant. Post-hoc tests, if relevant, were carried out by multiple comparisons with Bonferroni correction.

For Experiment 2, we calculated judgement probability for each speaking accent and skin stretch stimulation. LMM analysis was applied as in Experiment 1. The fixed factors were the Accent condition (S_a0, S_a1 and S_a2) and the Stimulation condition (P_a0, P_a1 and P_a2) and the random factor was the Participant.

3. RESULTS

Figure 2 shows relative judgement probability across the timing of somatosensory onsets in Experiment 1. It appears that lexical perception related to word segmentation changes depending on the timing of somatosensory stimulation. The percentage of judgement probability was reduced when the somatosensory stimulation led the first vowel (= P_t3), and was increased when somatosensory stimulation was lagged, more or less corresponding to the second vowel (= P_t6). LMM analysis showed a significant difference between stimulation conditions (χ^2 (7) = 31.26, $p < 0.01$). Post-hoc test showed that the amplitude of P_t3 was significantly smaller than the amplitudes of P_t2, P_t5, P_t6, P_t7 and P_t8 ($p < 0.04$ in all cases). Hence it appears that judgement probability depends on the timing of facial skin stretch stimulation.

Figure 3 represents judgement probability for the three speaking accent conditions (S_a0, S_a1 and S_a2) in Experiment 2. We found a reliable difference between auditory conditions (χ^2 (2) = 179.97, $p < 0.01$). Post-hoc test showed $p < 0.01$ in all combinations of speaking accent conditions. This suggests that a different accent in the phrase biases the word segmentation to extract lexical information as in [11]. We did not find any differences between somatosensory conditions (χ^2 (2) = 0.06, $p > 0.97$) and interaction effect between somatosensory and speaking accent condition (χ^2 (4) = 0.39, $p > 0.98$).

Figure 2: Judgement probability relative to P_t0 in Experiment 1. The error bars are standard error across participants.

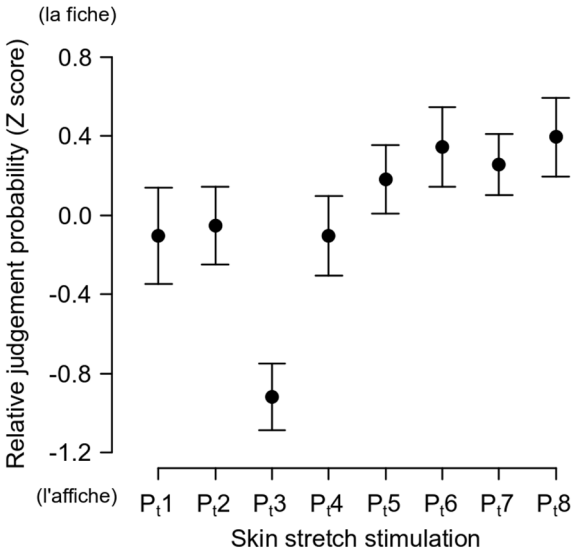
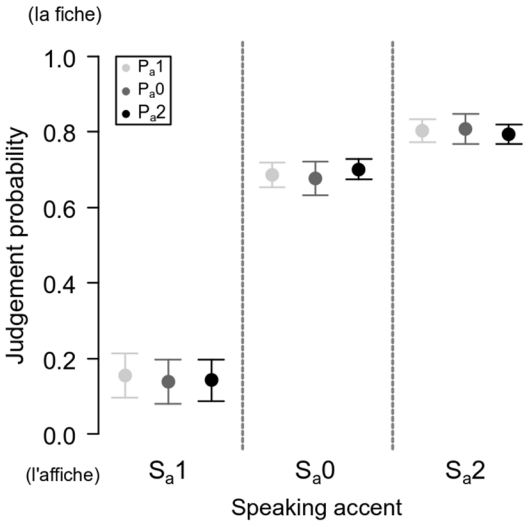


Figure 3: Judgement probability for three speaking accent conditions in Experiment 2. The error bars are standard error across participants.



4. DISCUSSION

A first and major finding of this study is that somatosensory inputs associated with facial skin deformation do modulate the perception of lexical information in French in relation with the timing of somatosensory stimulation relative to the target vowel in auditory stimuli. A second finding is that different amplitudes of somatosensory stimulation applied in coordination with the two driving vowels in the auditory stimuli do not modify lexical perception. These results demonstrate and quantify the capacity of orofacial somatosensory system to intervene in lexical perception associated to word segmentation. Given that similar somatosensory effects have been demonstrated in the perception of syllables (consonant-vowel sequences in [1]) and phonemes (American English vowels in [5]), our results extend the validity of somatosensory effects to the lexical level of speech processing, which required more complex (and/or higher-level) processing.

Since orofacial cutaneous mechanoreceptors provide kinesthetic information for speech production [4,6], the effect can be induced by somatosensory inputs which are expected to be accompanied in our own speech. This is consistent with studies in audio-visual speech perception. Indeed, it is well known that the intelligibility of speech sounds can be improved by providing visual information of the speaking movements in addition to the audio stimulus, in normal environments [8,11] as well as in noisy environment situations (e.g. [2,7,12]). Both visual and somatosensory inputs associated to speech-related movements hence appear to help to segment a speech phrase in ambiguous situations.

Lexical perception in our experiments was biased depending on when the somatosensory stimulation was applied. When somatosensory stimulation was applied before the first vowel presentation, the perception was biased toward “affiche”. On the other hand, when the somatosensory stimulation was applied between the first and second vowel, the perception was rather biased toward “fiche”. In both cases, when the somatosensory input preceded one specific vowel, the perception was biased toward the vowel. This is consistent with the finding that the change of cortical potentials by auditory-somatosensory interaction was induced specifically when somatosensory inputs precede auditory inputs [3]. Since anticipatory articulatory movements precede speech sounds in speech production, this temporal relationship between somatosensory and auditory inputs might be important to induce the somatosensory effect for speech perception.

We did not find any effect of somatosensory amplitude in Experiment 2. This may be due to over-simplification of somatosensory stimulation. We expected that amplitude difference in somatosensory stimulation would produce different sensations of hyper-articulation associated to the current auditory stimuli. The first vowel in our audio stimuli was always /a/ (l’a or la). On the other hand, the second vowel did vary over the 17 words (8 low-, 7 mid-, and 2 high-vowels). Since the pattern of skin deformation can be different between low-, mid- and high-vowels [13], somatosensory inputs that would be received in speech production should be different in the first and second vowels. This suggests that the current manipulation may not represent correctly the actual situation of somatosensory inputs arising from speech movement corresponding to the current auditory stimuli. More complicated and realistic patterns of stimulation might be required to induce an effect in this paradigm.

5. ACKNOWLEDGEMENTS

This work was supported by the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 339152). We also thank Nathan Mary and Dorian Deliquet for data collection and Silvain Gerber for statistical analysis.

6. REFERENCES

[1] Gick, B., Derrick, D. 2009. Aero-tactile integration in speech perception. *Nature*, 462:502–504.

[2] Grant, KW., Seitz, P-F. 2000. The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am*, 108(3)(Pt. 1):1197–1208.

[3] Ito, T., Gracco, VL., Ostry, DJ. 2014. Temporal factors affecting somatosensory-auditory interactions in speech processing. *Front Psychol*, Available from: 10.3389/fpsyg.2014.01198.

[4] Ito, T., Ostry, DJ. 2010. Somatosensory Contribution to Motor Learning Due to Facial Skin Deformation. *J Neurophysiol*, 104(3):1230–1238.

[5] Ito, T., Tiede, M., Ostry, DJ. 2009. Somatosensory function in speech perception. *Proc Natl Acad Sci U S A*, 106(4):1245–1248.

[6] Johansson, RS., Trulsson, M., Olsson, KÅ., Abbs, JH. 1988. Mechanoreceptive afferent activity in the infraorbital nerve in man during speech and chewing movements. *Exp Brain Res*, 72:209–214.

[7] Kim, J., Davis, C. 2004. Investigating the audio-visual speech detection advantage. *Speech Commun*, 44:19–30.

[8] Munhall, KG., Jones, JA., Callan, DE., Kuratate, T., Vatikiotis-Bateson, E. 2004. Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception. *Psychol Sci*, 15(2):133–137.

[9] R Core Team. 2019. R: A language and environment

for statistical computing. [Internet]. Vienna, Austria, R Foundation for Statistical Computing, Available from: https://www.r-project.org/.

[10] Schwartz, J-L., Basirat, A., Ménard, L., Sato, M. 2012. The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *J Neurolinguistics*, 25:336–354.

[11] Strauss, A., Savariaux, C., Kandel, S., Schwartz, J-L. 2015. Visual lip information supports auditory word segmentation. *FAAVSP 2015*, Vienna, Austria.

[12] Sumby, WH., Pollack, I. 1954. Visual Contribution to Speech Intelligibility in Noise. *J Acoust Soc Am*, 26(2):212–215.

[13] Vatikiotis-Bateson, E., Kuratate, T., Kamachi, M., Yehia, H. 1999. Facial deformation parameters for audiovisual synthesis. *AVSP'99*, Santa Cruz, CA, USA, 118–122.