



HAL
open science

Orofacial somatosensory inputs improve speech sound detection in noisy environments

Rintaro Ogane, Jean-Luc Schwartz, Takayuki Ito

► **To cite this version:**

Rintaro Ogane, Jean-Luc Schwartz, Takayuki Ito. Orofacial somatosensory inputs improve speech sound detection in noisy environments. SNL 2018 - 10th Annual Society for the Neurobiology of Language Conference, Aug 2018, Québec City, Canada. hal-01859597

HAL Id: hal-01859597

<https://hal.science/hal-01859597>

Submitted on 29 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Orofacial somatosensory inputs improve speech sound detection in noisy environments

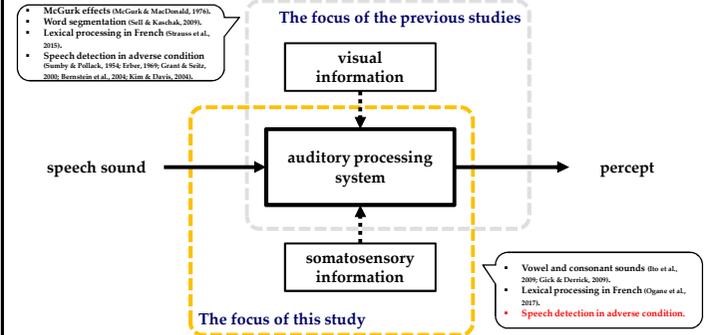
Rintaro Ogane¹, Jean-Luc Schwartz¹, Takayuki Ito^{1,2}

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

Contact
rintaro.ogane@gipsa-lab.grenoble-inp.fr

Introduction

- Speech perception is an interactive process with multiple modes and probably some perceptuo(multisensory)-motor connections (Schwartz et al., 2012).

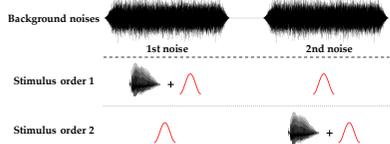


- The aim of this study.
 - To investigate a functional role of **somatosensory** information on the perception of speech.
- The hypothesis of the current experiment.
 - Somatosensory information associated with facial skin deformation facilitates correct detection of speech sounds in low signal to noise ratio (SNR) environment.

Methods

- We carried out a speech detection test in noise (10 participants) under 2 conditions, (1) with somatosensory information condition (**SKIN**) and (2) no somatosensory information (audio-only) condition (CTL).
- The subject task was to identify which noise period includes the target utterance.
- Target utterance is /pa/ recorded by a native French speaker.
- 8 SNR levels (from -8 to -15 dB) between target utterance and background noises (white noises) were tested.
- 6 Hz of a sinusoidal pattern was used as the skin stretch stimulation to mimic the speaking motion.

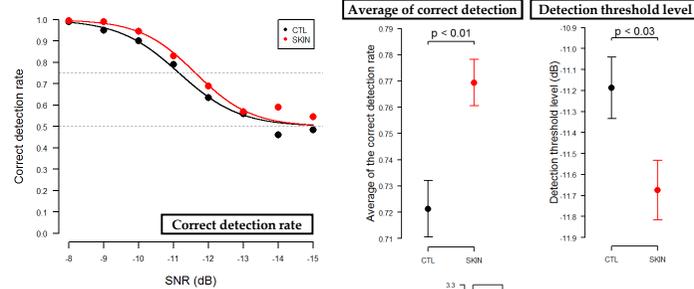
The settings for the **SKIN** condition.



Results

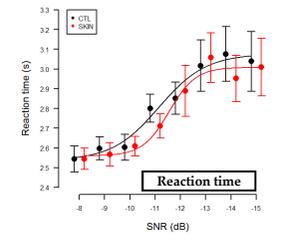
Correct detection rate & detection threshold level

- The **correct detection rate** for the **SKIN** condition was significantly greater than for the CTL condition ($t_{(9)} = 3.83, p < 0.01$).
- The **detection threshold level** for the **SKIN** condition was significantly smaller than for the CTL condition ($t_{(9)} = -2.29, p < 0.03$). → **“audio-somatosensory detection advantage”**.



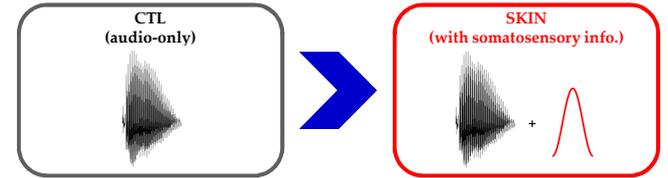
Reaction time

- In both conditions, the reaction time at lower SNRs was longer than at higher SNRs ($F_{(7, 63)} = 14.53, p < 0.01$).
- There was no significant difference in the experiment condition ($F_{(1, 9)} = 1.07, p > 0.3$), nor an interaction effect ($F_{(7, 63)} = 1.04, p > 0.4$).



Discussion

- The **detection threshold level** for CTL and **SKIN** conditions:



This result is consistent with the audio-visual speech perception literature.

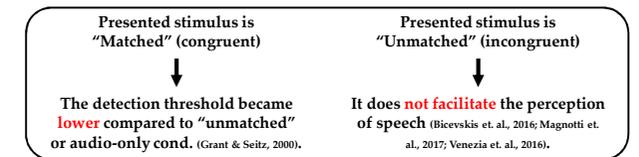
- The speaker's facial movements enhance the perception of speech (Sumbly & Pollack, 1954).
- Lipreading in noise enhances the detection of speech (Bernstein et al., 2004; Schwartz et al., 2004; Tjan et al., 2014).

The reduction of **detection threshold level** → 0.49 dB in average.

- cf. audio-visual speech perception.
 - The mean threshold level was lowered about 3 dB (Bernstein et al., 2004), 1.56 dB (Grant & Seitz, 2000).

This suggests that somatosensory information also provides a rich source to detect the embedded target utterance in noisy environments as well as visual information.

- The relationship between the timings of somatosensory stimulation and output of the target speech sound.



- Our result is consistent with the one in “Matched” condition.
- This suggests that timing of stimulus is important to detect the speech utterance in noise, and we expect that the facilitation would not arise in the unmatched condition.

Summary

- Somatosensory inputs facilitated the correct detection of speech sounds in low SNR environments compared to audio-only condition.
- This result suggests that somatosensory information can intervene in the speech detection process in noisy environments.

Acknowledgments

- This research was supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 339152, “Speech Unit(es)”).
- We wish to thank Gurvan Quiniou for his help in data collection and analysis.

