Visual lip information supports auditory word segmentation
Antje Strauß, Christophe Savariaux, Sonia Kandel, Jean-Luc Schwartz

To cite this version:
Antje Strauß, Christophe Savariaux, Sonia Kandel, Jean-Luc Schwartz. Visual lip information supports auditory word segmentation. FAAVSP 2015, Sep 2015, Vienna, Austria. <halshs-01298351>

HAL Id: halshs-01298351
https://halshs.archives-ouvertes.fr/halshs-01298351
Submitted on 5 Apr 2016

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.
Visual lip information supports auditory word segmentation

Antje Strauß, Christophe Savariaux, Sonia Kandel, Jean-Luc Schwartz
CNRS, GIPSA-Lab, F-38000 Grenoble, France; Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France
antje.strauss@gipsa-lab.fr

Introduction

Speech segmentation has been shown to depend on statistical learning of auditory regularities, e.g. transitional probabilities between syllables [1] and prosodic cues including fluctuations in intensity, F0, segment durations and various articulatory components [2,3]. However, up to now benefits of visual prosodic cues for speech segmentation have only been investigated in artificial languages [4,5]. We hypothesize that lip information are used in natural speech when word segmentation is difficult as for example in the case of liaisons in French.

Methods

- 17 French sentences were created consisting of the carrier phrase “C’est” [eng. “That is”] followed by a determiner and a noun that allowed two possible readings either with liaison (e.g., “l’amarre” [the rope] (A1)) or without liaison (e.g., “la mare” [the pond] (A2)).
- The speaker was instructed to produce 10 auditorily hyper-articulated repetitions of each possible reading (A1, A2) and 10 ambiguous utterances (Axx).
- Afterwards, the speaker was listening to his recordings while producing lip movements in synchrony: 5 times visually hyper-articulating each possible reading (V1, V2).
- Audiospeech stimuli were recorded using a PAL camera (SONY HDR-XR500E) with a sampling rate of 25 images per second and an AKG (C-1005) microphone for the audio track.
- Lips were colored in blue to be able to apply a chromakey on each image leaving only lip contours. Lip parameters (width, height, surface) were extracted by using the Tacle software developed at Gipsa [6].

Results

- Figure 3: Analysis of stimulus characteristics separated for 2- and 3- or 4-syllabic words. A ratio between the two vowels of interest (difference divided by sum) is calculated. Positive values indicate that F0 (A) or lip surface (B) are higher in the second compared to the first vowel of interest. A, F0 dissociates A1 and A2 but not ambiguous stimuli. B, Lip surface dissociates V1 and V2.
- Behavioural results.

Discussion

The current data suggest that visual lip information could have an impact on word segmentation processes. This is particularly relevant for ambiguous utterances like the ones that embed liaisons in French.

1. Lip movements hyper-articulating the CV-segmentation (V1, e.g. “la mare”, in opposition to the VC-segmentation V1, e.g. “Tamara”) increases decisions for the CV-segmentation in all audiovisual conditions compared to audio-only.
2. Contrary to our prediction, this response gain by lip information was the same even if acoustic cues were ambiguous.

Our data are in line with studies that show an influence of visual lip information when listening to speech. The results extend evidence showing the usage of lip information in contrastive prosody [7], during multi-stable speech perception [8], and to trigger lexical access [9]. Thus, visual speech does not only provide segmental but also suprasegmental and prosodic cues which enable the perceiver to successfully segment words from a continuous speech stream.

References


This research is funded by the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 339152, “Speech Unit(s)”).