Electrophysiological evidence for audio-visuo-lingual speech integration

Coriandre Vilain, Avril Treille, Marc Sato

ELECTROPHYSIOLOGICAL EVIDENCE FOR AUDIO-VISUO-LINGUAL SPEECH INTEGRATION

Coriandre Vilain[1], Avril Treille[1] & Marc Sato[2]

[1] *GIPSA-lab, Département Parole & Cognition, CNRS & Grenoble Université, Grenoble, France*
[2]*Laboratoire Parole & Langage, CNRS & Aix Marseille Université, France*

Introduction

Audio-visual speech perception is a special case of multisensory processing that interfaces with the linguistic system. One important issue is whether cross-modal interactions only depend on well-known auditory and visuo-facial modalities or, rather, might also be triggered by other sensory sources less common in speech communication. The present EEG study aimed at investigating cross-modal interactions not only between auditory, visuo-facial and audio-visuo-facial syllables but also between auditory, visuo-lingual and audio-visuo-lingual syllables.

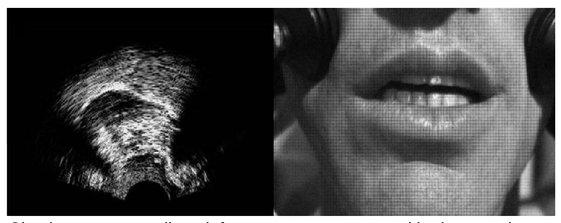
METHODS

❖ Participants

18 healthy adults, right-handed native French speakers.

❖ Stimuli

- **/pa/**, **/ta/** and **/ka/** syllables were individually recorded by one male and one female speakers in a sound-proof room. Two **/pa/**, **/ta/** and **/ka/** tokens were selected per speaker .
- Synchronous recordings of auditory, visual and ultrasound lingual signals were acquired by a Terason T3000 ultrasound system **(UltraSpeech system, Hueber et al. 2008)** .



Simultaneous recording : left : tongue movements with ultrasound system; right : Lips movements with camera

- 60 stimuli were created consisting of 12 /pa/, /ta/ and /ka/ syllables related to 5 conditions: an auditory condition (**A**), 2 visual (**VL**, **VT**) and 2 audio-visual (**AVL**, **AVT**) conditions related to either lip or tongue movements of a speaker.

❖ **Method**

1) Before the experiment, short explanations on the ultrasound system and tongue movements required for the production of **/pa/**, **/ta/** and **/ka/** syllables were given.

2) Behavioral session: a three-alternative forced-choice identification task, with participants instructed to categorize as quickly as possible each perceived syllable with their right hand.

3) EEG session: a three-alternative forced-choice identification task, with participants instructed to categorize each perceived syllable with their right hand, after an audio "beep".

❖ Data acquisition

- EEG data were continuously recorded from 64 scalp electrodes (international 10–20 system) using the Biosemi ActiveTwo AD-box EEG system operating at a sampling rate of 256 Hz.
- Two additional electrodes served as reference [CMS] & [DRL]
- One other external reference electrode was at the top of the nose. The electrooculogram controlling for horizontal (HEOG) and vertical (VEOG) eye movements were recorded using electrodes at the outer canthus of each eye as well as above and below the right eye. Before the experiment, the impedance of each electrode was adjusted to get low offset voltage and stable DC.

❖ Analysis

- *Behavioral analyses (1):*

	Behavioral experiment	EEG experiment
%	ANOVA : Environment (**Noise/No Noise**), Articulator (**Lips/Tongue**), Modality (**A, AV, V**)	ANOVA : Environment (**Noise/No Noise**), Articulator (**Lips/Tongue**), Modality (**A, AV, V**)
RT's	ANOVA : Environment (**Noise/No Noise**), Articulator (**Lips/Tongue**), Modality (**A, AV, V**)	

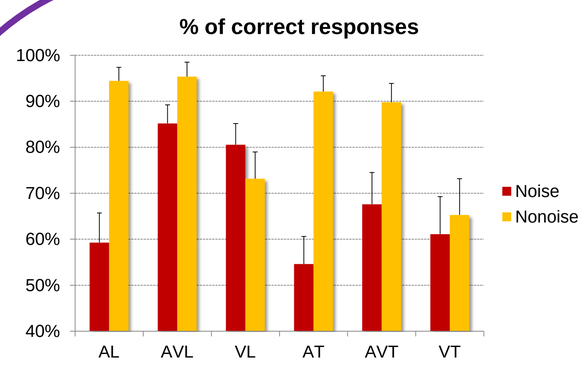
- *EEG analyses on fronto-central electrodes (F3/F4/C3/C4/Fz/Cz):*

Pre-processing	- Re-referenced off-line to the nose - Epochs : 1000ms (baseline from -500 to -400ms to the acoustic syllable onset)	- Filtering : 1-30 Hz - Rejection : ±60 µV

➔ 1) N1 and P2 **latency** and **amplitude** analysis by ANOVA with Environment (**Noise/No noise**), Articulator (**Lips/Tongue**) and Modality (**A, AV**) as intra-subject variable (main effects and interactions)
➔ 2) N1 and P2 latency and amplitude were also tested for **integration**, i.e. an ANOVA with Environment (Noise/No noise), Articulator (Lips/Tongue), Signal type (**Bimodal (AV)/Sum (A+V)**) as intra-subject variable (main effects and interactions)

RESULTS

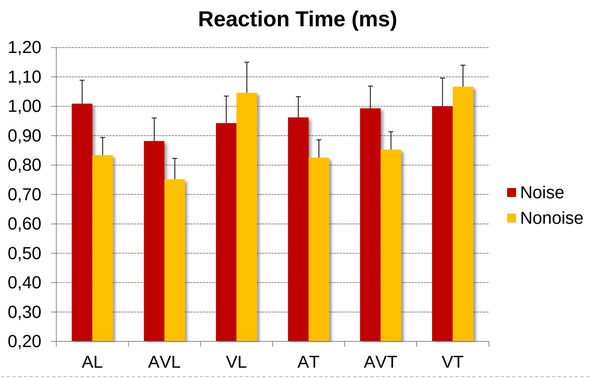
1) Behavioral results



% Correct Response :

Main Effects: - Noise < No noise (p<.0001)
- Lips > tongue (p<.005)
- AV > A = V (p<.0001)

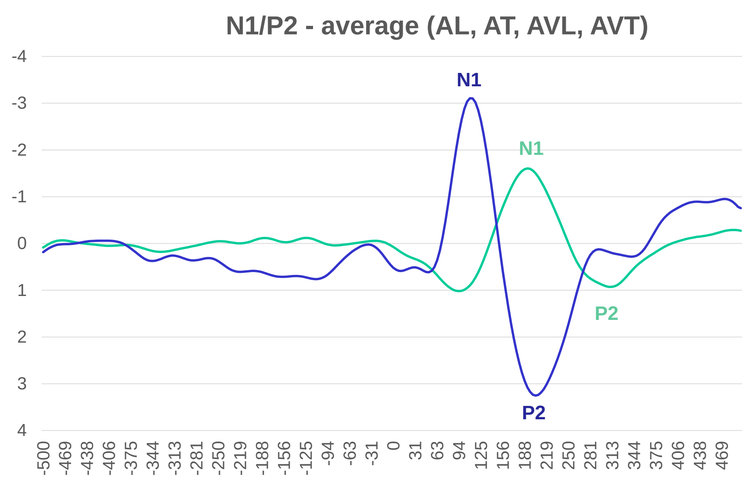
Interactions: - Noise: AV = V > A
No noise: AV=A > V (p<. 0001)
- Lips: A=V < AV
Tongue: AV=A > V (p<.05)



TR :

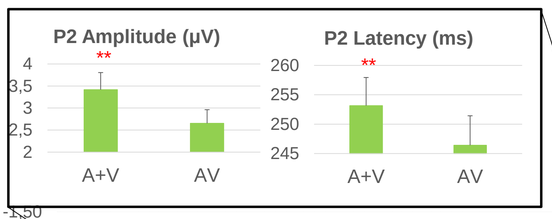
Main Effects: - Noise > No Noise (p<.03)
- A=AV < V (p<.0001)
Interactions: - Noise: A = AV = V
No noise: A = AV < V (p<.0001)
- Lips: A = V > AV
Tongue: A = AV < V (p<.01)

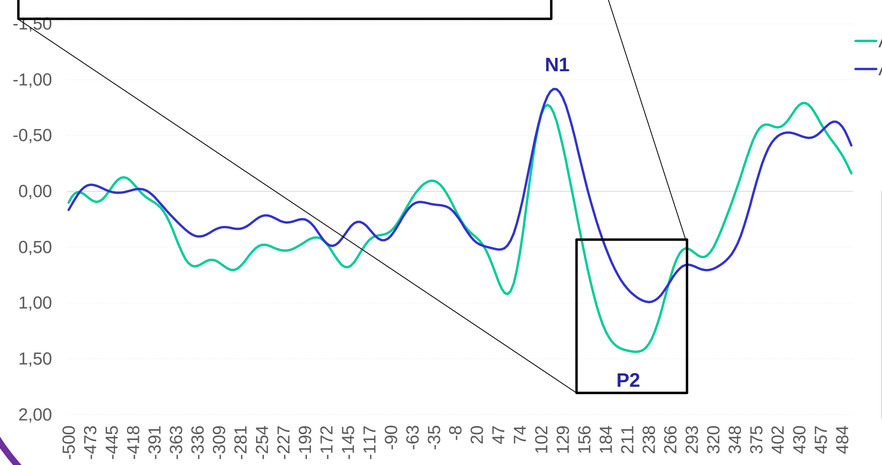
2) EEG – *Noise vs No Noise*



Noise causes:
- **Reduced amplitude** (p<.005)
- **Delayed Latency** (p<.0001)

3) EEG – *Integration (AV <> A+V)*

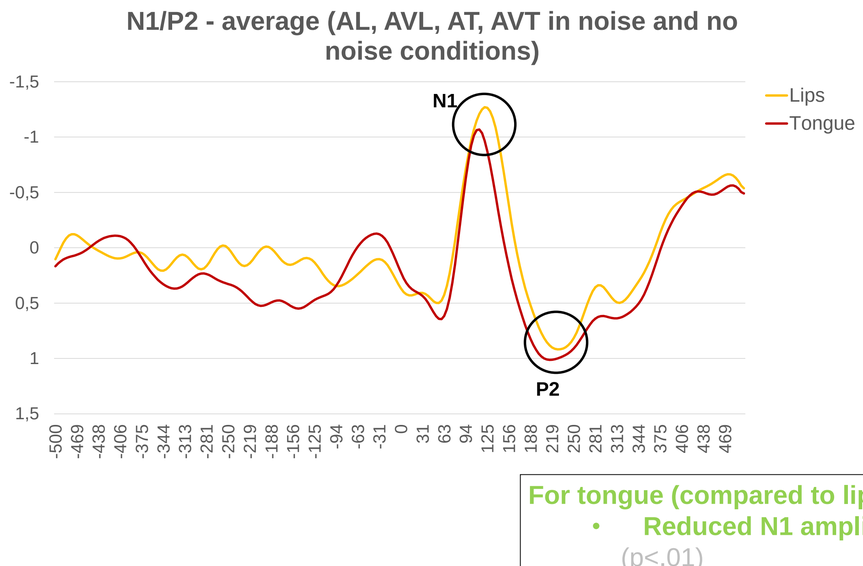




N1 : No effect
P2 : Reduced amplitude (p<.0001) **and latency** (p<.007) **in AV compared to A+V for both lip and tongue signals**

Amplitude P2 : A+V : Lips (2,91uV); Tongue (2,36uV)
AV : Lips (3,92uV); Tongue (2,94uV)
Latency P2 : A+V : Lips (258ms); Tongue (248ms)
AV : Lips (248ms); Tongue (244ms)

4) EEG – *Lips vs Tongue*



For tongue (compared to lips) :
- **Reduced N1 amplitude** (p<.01)
- **Larger P2 amplitude** (p<.0001)

Discussion

1) **Lips > tongue > chance** : For the visual-only modality, syllables are perceived correctly even when they refer to tongue movement.
2) **Noise vs No Noise** : In line with previous studies, adding noise reduces and shifts the N1/P2 peaks.
3) **Integration** : Early integration processing (on P2) during AV-lips and AVT-tongue speech perception compared to A+V.
4) **Difference between lips and tongue audio-visual speech perception** : an amplitude difference (N1 is smaller and P2 is larger than N1/P2 observed during AV-lips perception)

=> Although participants were not experienced with visuo-lingual stimuli, our results demonstrate that they were able to recognize them and provide the first evidence for audio-visuo-lingual speech interactions. These results further emphasize the multimodal nature of speech perception and likely reflect the impact of listener's knowledge of speech production on speech perception.

Part of this research was supported by a grant from the European Research Council (FP7/2007-2013 Grant Agreement no. 339152, "Speech Unit(e)s")
Correspondence: avril.treille@gipsa-lab.inpg.fr