The role of acoustic cues in speech perception during homophonic segmentation

Maria del Mar Cordero Rull

Université Côte d'Azur, CNRS, BCL,

Direction : Fanny Meunier

segmentingwordsfromacontiuousstreamisnotatr
ivialtaskatallespeciallyifyouarenotfamiliarwitht
helanguageyouareexposedtoinyourenvironment

The problem of the continuity of acoustic speech signal

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

UNIVERSITÉ CÔTE D'AZUR

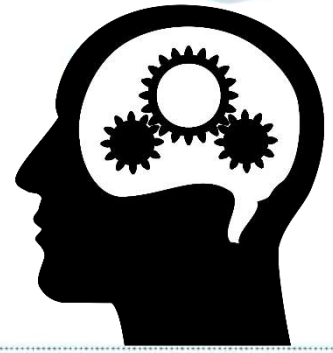
BASES CORPUS LANGAGE

cnrs

UNIVERSITÉ Grenoble Alpes

CoCoLab
COMPLEXITY AND COGNITION LAB

segmentingwordsfromacontiuousstreamisnotatr
ivialtaskatallespeciallyifyouarenotfamiliarwitht
helanguageyouareexposedtoinyourenvironment



RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

UNIVERSITÉ CÔTE D'AZUR

BASES CORPUS LANGAGE

cnrs

UNIVERSITÉ Grenoble Alpes

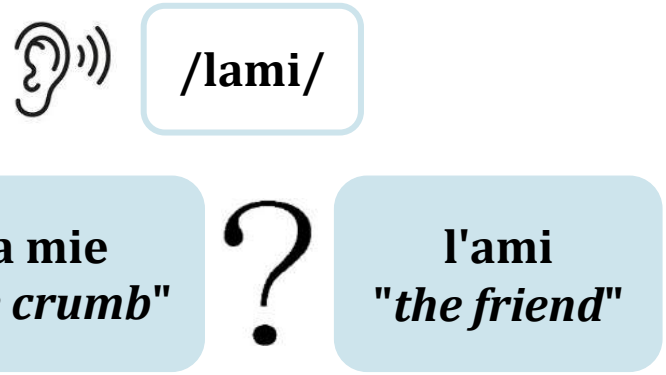
CoCoLab
COMPLEXITY AND COGNITION LAB

The problem of speech segmentation

- Natural speech is highly complex and variable
- Linguistic input enable to activate or inhibit word representations in mental lexicon (e.g., Gaskell & Marslen-Wilson, 1997)
- Listeners use L1 tacit knowledge to achieve speech segmentation (e.g., Quené, 1992)
- Lexical competition → solve ambiguities and recognise target meaning (Norris, 1994; McClelland and Elman, 1986)

The problem of speech segmentation

- Daily transient segmentation ambiguities
- **What happens to homophonic sequences?**
- Correct segmentation must consider fine-grained acoustic details in speech signal



Exploitation of speech cues

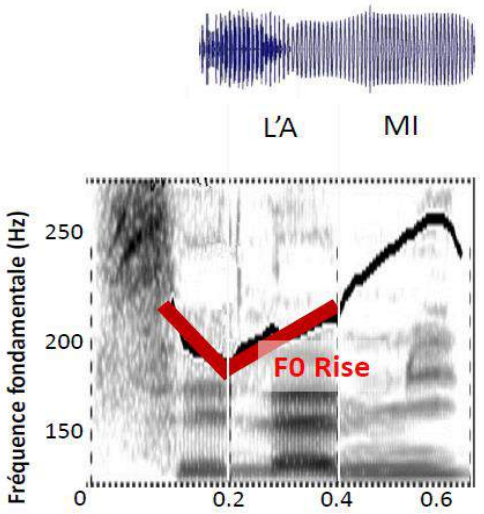
- Native listeners solve speech segmentation by relying on acoustic-phonetic, phonological, lexical, rhythmic, and statistical regularities (Cutler & Norris, 1988; Mattys et al., 2005; Saffran et al., 1996)
- In French → challenge of resyllabification
- If lexical information is informative > acoustic-phonetic cues (Spinelli et al., 2002)
- If lexical knowledge is uninformative < acoustic-phonetic cues
  e.g., *les ailes* vs. *les zèles*, both [le. zɛl] (Shoemaker, 2014; Spinelli et al., 2003)

Exploitation of speech cues in French I

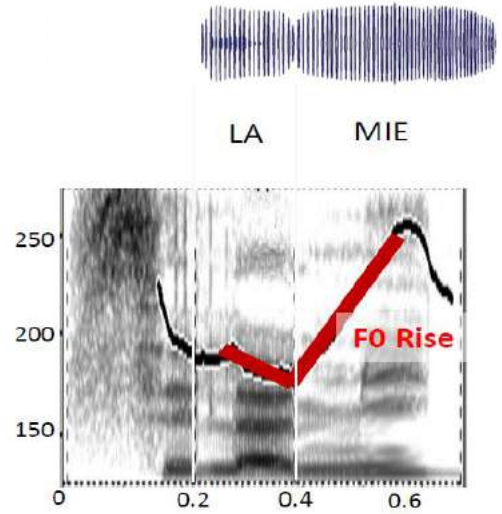
Homophonic sequences like [sɛlami]

- F0 rise often found at the beginning of content words
(Welby, 2003a, 2003b, 2007)

- Can discriminate by 66.3% and identify it correctly by 75.5% (Spinelli et al., 2007)



L'ami = "The friend"



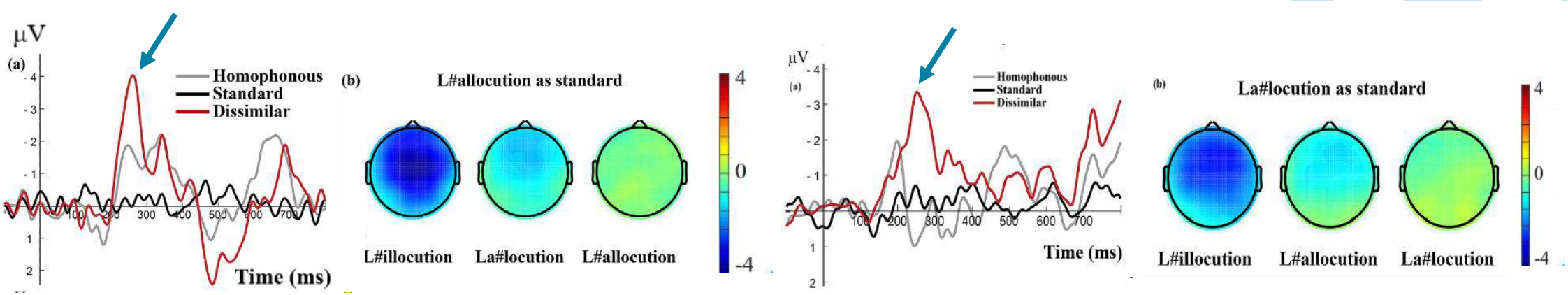
La mie = "The crumb"

Acoustic information in /lami/ is used when presented in isolation **(offline)**

Exploitation of speech cues in French II

Homophonic sequences like [sɛlami]

- Can discriminate by 79%
  (Do Carmo-Blanco et al., 2019)



Acoustic information in /lami/ is used when presented in isolation **(online)**

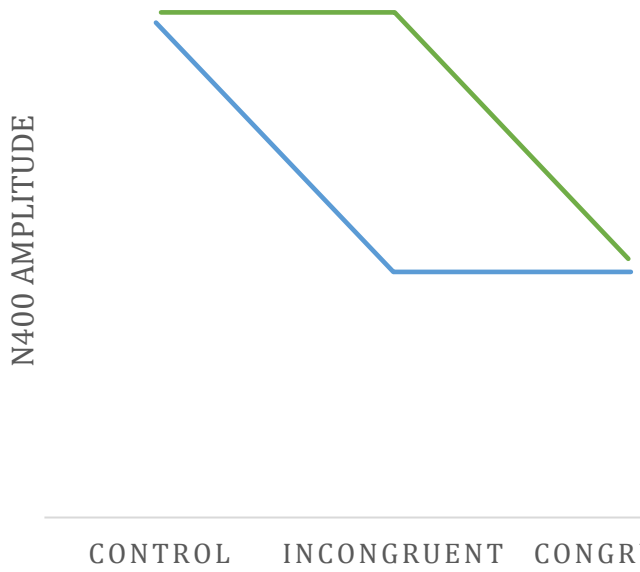
ERPs: light to online speech processing

- Right centroparietal negative component peaking at ~400 ms
  (e.g., Kutas and Hillyard, 1984)
- Ease with which a word or a sentence is integrated into a given context (Chwilla et al., 1995; Friederici, 1995)
  - Incongruent sentences elicit larger negativity than congruent sentences

(Kutas & Federmeier, 2011)

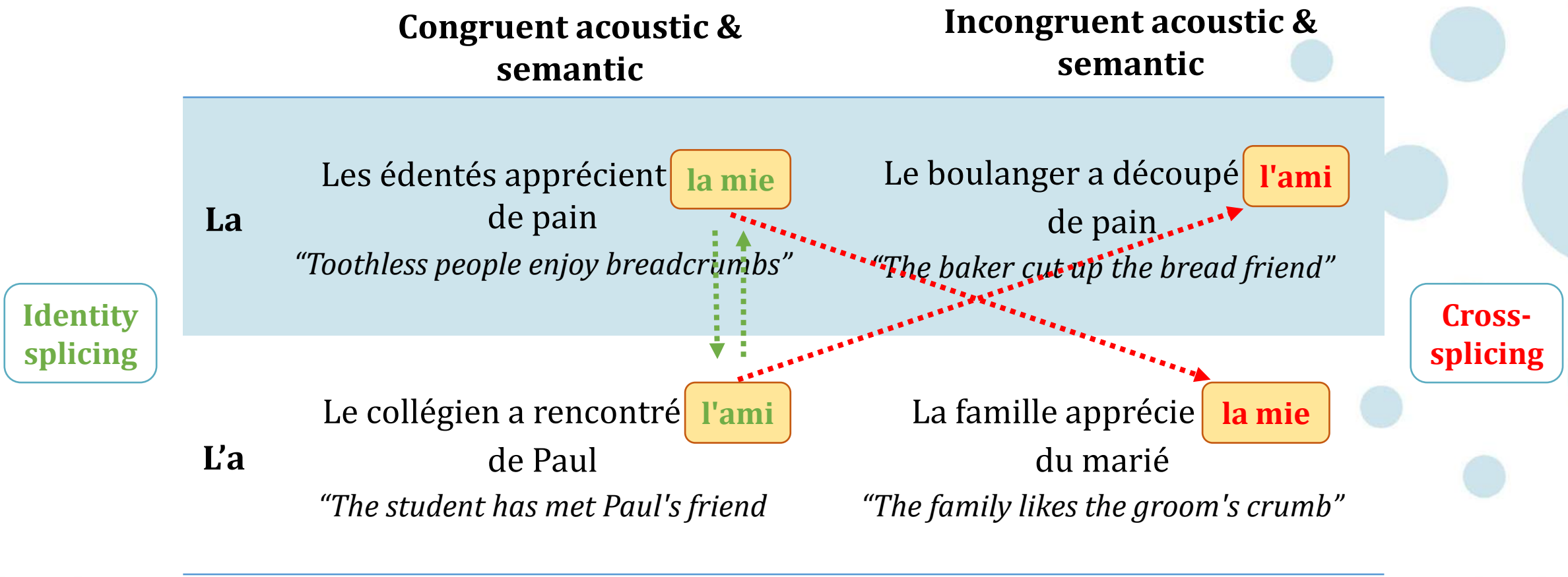
Interest of study and main hypotheses

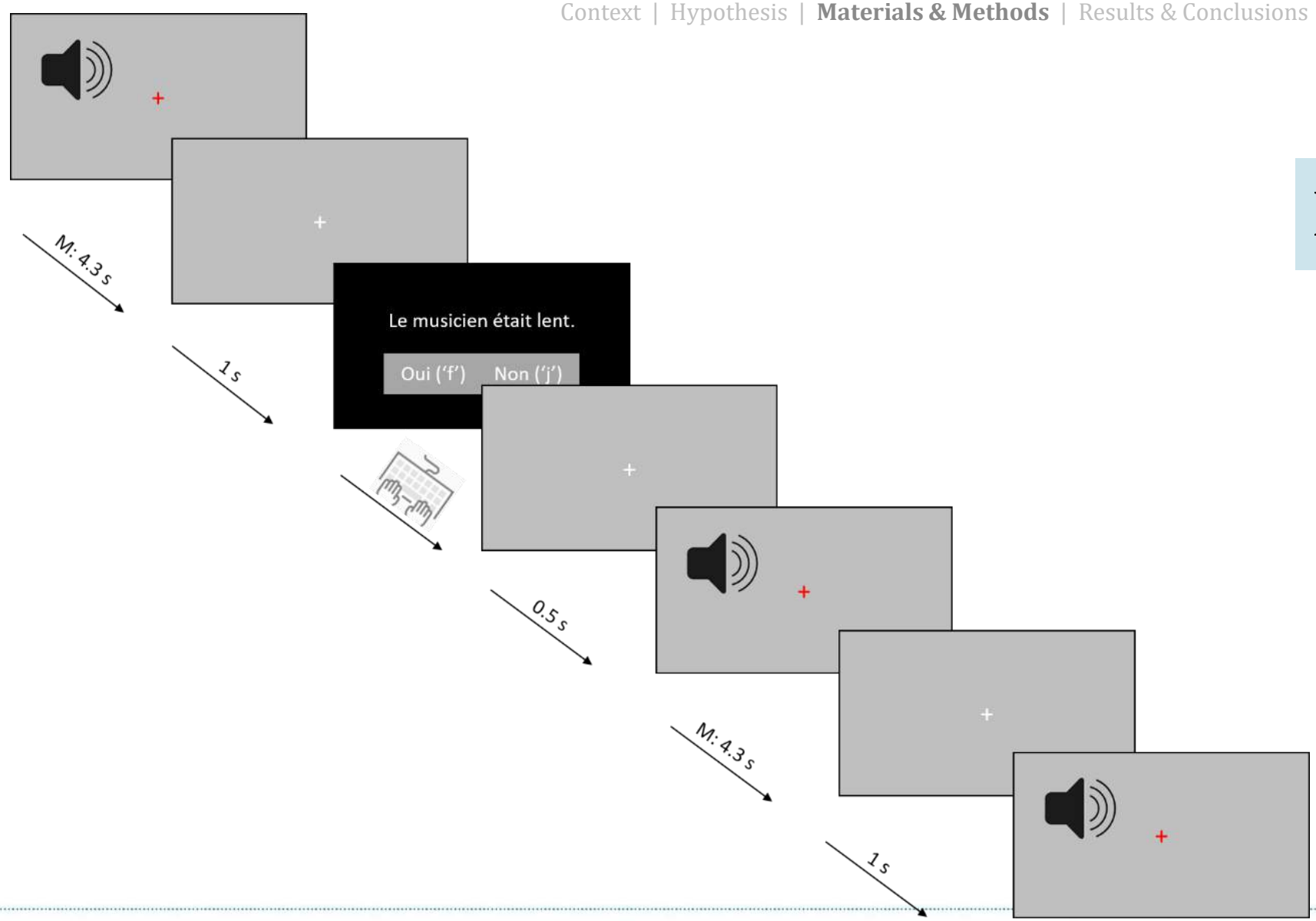
- Examine cues involved in processing word boundaries in French
- How does the sensitivity to fine-grained acoustic details change at sentence level?



- Hyp. 1: Fine acoustic analysis occurs in real-time speech processing
- Hyp. 2: Homophonic representations are accessed through contextual means

Stimuli


Congruent acoustic & semantic
Incongruent acoustic & semantic
La
Les édentés apprécient la mie de pain
"Toothless people enjoy breadcrumbs"
Le boulanger a découpé l'ami de pain
"The baker cut up the bread friend"
L'a
Le collégien a rencontré l'ami de Paul
"The student has met Paul's friend
La famille apprécie la mie du marié
"The family likes the groom's crumb"
Identity splicing
Cross-splicing



Experimental Design

EXPERIMENT 1: EGG

No active task

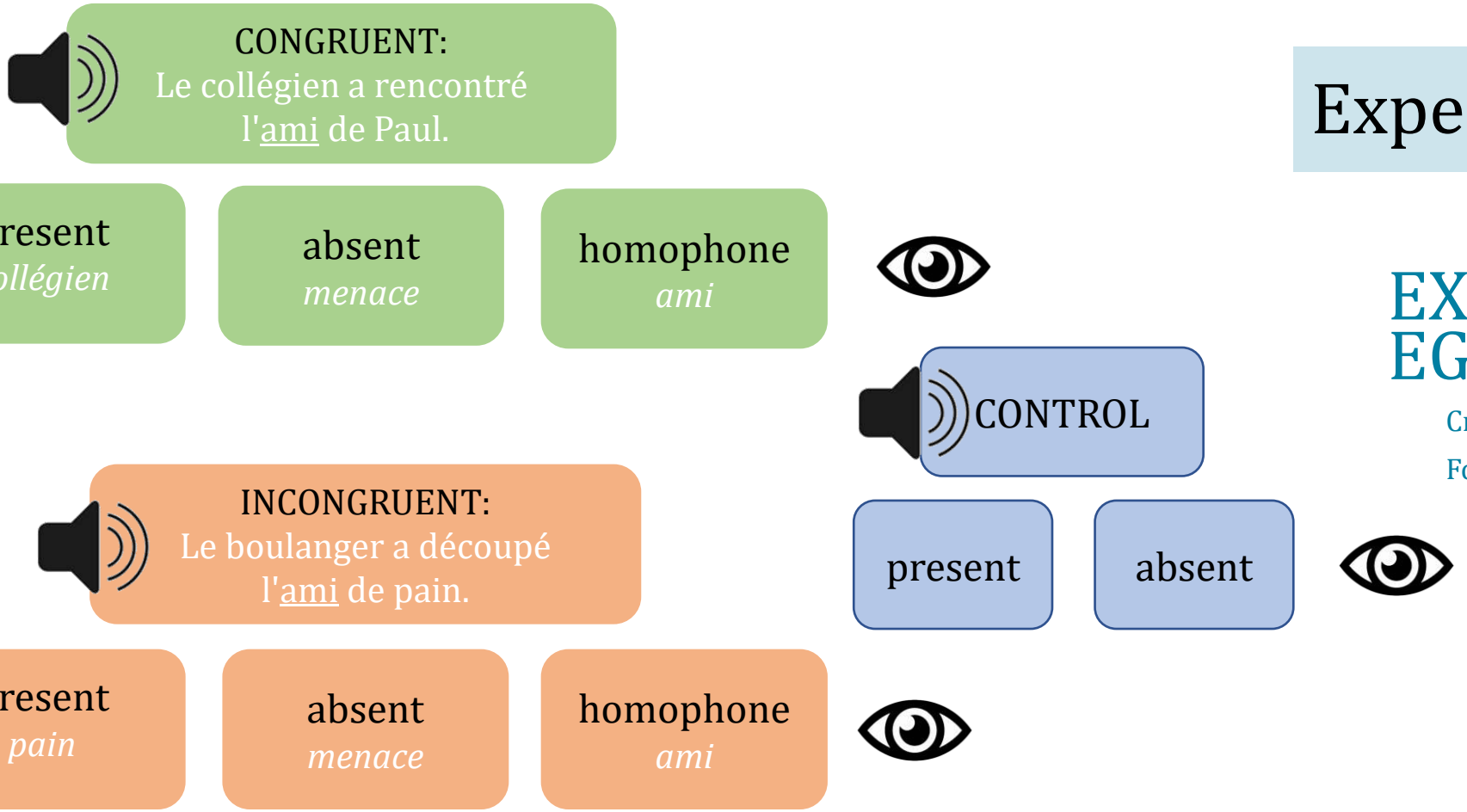
Semantic **judgement task** as control for attention

Experimental Design

EXPERIMENT 2: EGG & Behavioural

Cross-modal **recognition task**

Focus on lexical information (words)



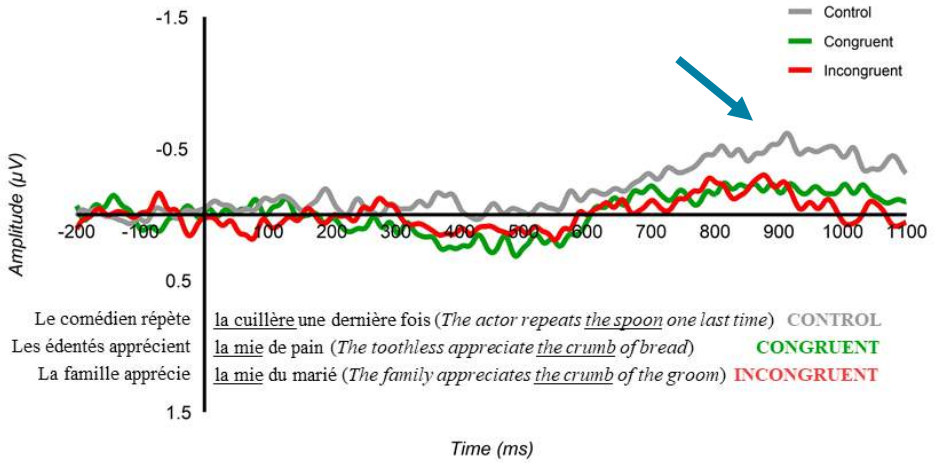
/ La famille apprécie
la mie du marié /

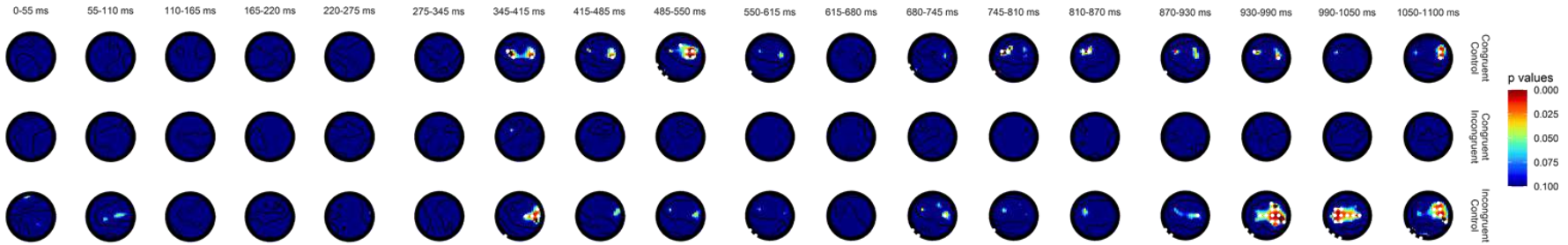
+

MIE

Results & Conclusions : EEG (online measures)

EXPERIMENT 1





→ Congruent = Incongruent in **Experiment 1**

- Fine-grained acoustic information NOT used as cue when exposed to informative-enough context
- When semantic information is present, listeners do not rely on acoustic cues to achieve comprehension

Congruent vs Control
- 345 – 550 ms
- 745 – 1100 ms

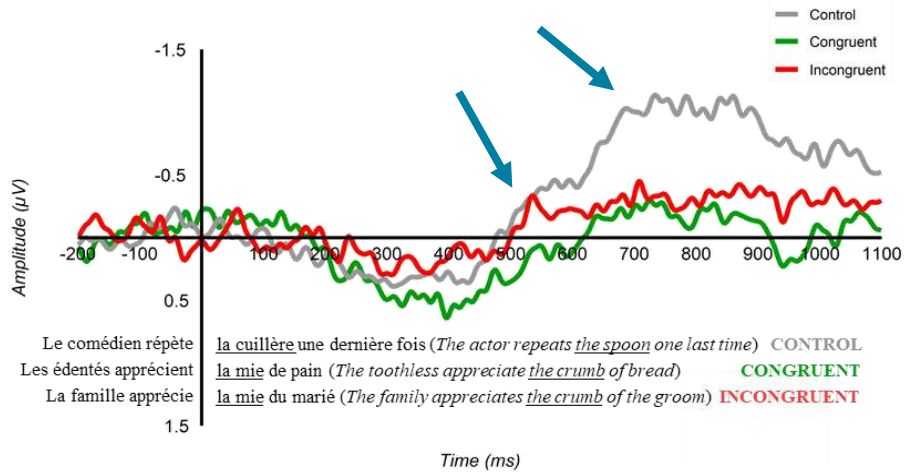
Incongruent vs. Control
- 345 – 415 ms
- 870 – 1100 ms

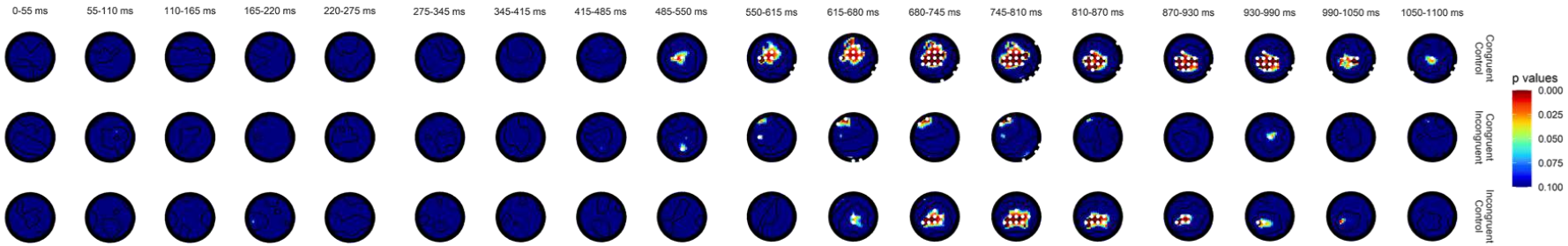
Congruent vs. Incongruent
- *ns*

Global Field Power (GFP): Skrandies, 1990; Delorme and Makeig, 2004; Permutation analyses: Maris & Oostenveld, 2007

Results & Conclusions : EEG (online measures)

EXPERIMENT 2





→ Congruent # Incongruent in **Experiment 2**

- When attention paid to lexical information → Listeners seem to perceive acoustic variations
- Use of cues seems to be situational-dependent
  (inconsistent with Mattys et al. 2005 & 2007)

Congruent vs Control
485 – 1100 ms

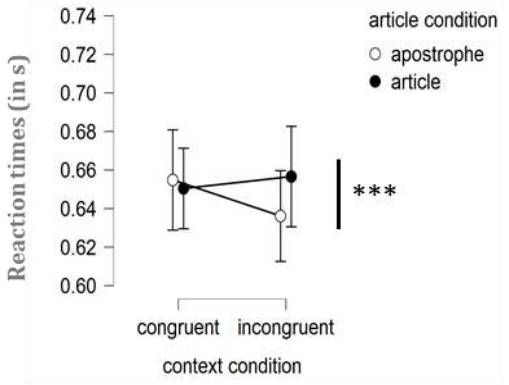
Incongruent vs. Control
615 – 990 ms

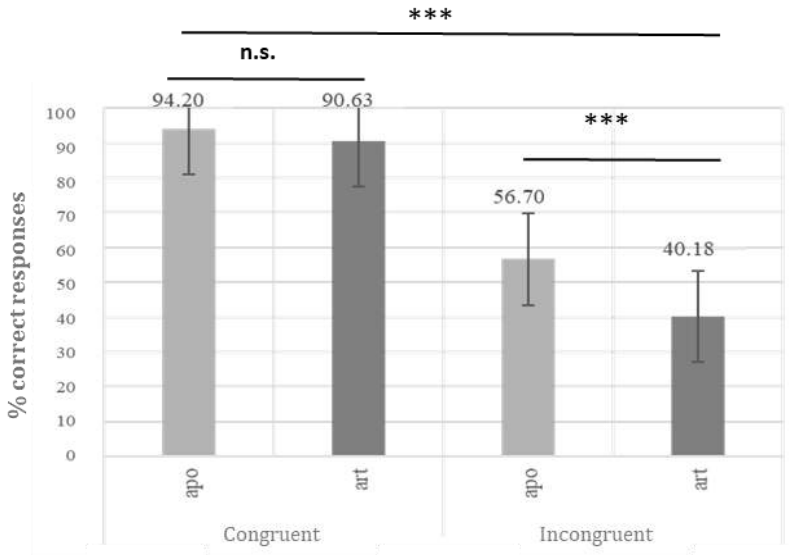
Congruent vs. Incongruent
615 – 810 ms

Global Field Power (GFP): Skrandies, 1990; Delorme and Makeig, 2004; Permutation analyses: Maris & Oostenveld, 2007

Results & Conclusions : Behavioural (offline measures)

EXPERIMENT 2





- Context plays a role when recognizing homophonic sequences
  - Incongruent contexts had lower accuracy rates & slower RTs
- Resyllabification does not seem to difficult word-boundary recognition in French
- Speech perception is modulated by variations in acoustic indices of the signal

References

Bell, A. J., & Sejnowski, T. J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation, 7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review, 81*(4), 348–374. https://doi.org/10.1037/h0036656

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001a). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language, 45*, 177-199.

Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. (2001b). Sequence detection in nonwords in French: Where is the syllable effect? *Language and Cognitive Processes, 16*, 609-636.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21.

Do Carmo-Blanco, N., Hoen, M., Pota, S., Elsa Spinelli, E., Meunier, F. 2019. Processing of non-contrastive subphonemic features in French homophonous utterances: An MMN study. *Journal of Neurolinguistics, 52*.

Kutas, M., Federmeier, K. D. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology, 62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48*(6), 609–621. https://doi.org/10.1016/0013-4694(80)90419-8

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Spinelli, E., Grimault, N., Meunier, F., Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Attention, Perception, & Psychophysics. 72*, 775-787. https://doi.org/10.3758/APP.72.3.775

Spinelli, E., Welby, P., Schaegis, A.-L. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of French elision. *Language and Cognitive Processes, 22*:6, 828-859.

Vroomen, J., de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 710-720.

Welby, P. (2003a). French intonational rises and their role in speech seg mentation. In *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology* (pp. 2125-2128). Geneva, Switzerland.

Welby, P. (2003b). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment; their role in speech segmentation*. PhD dissertation, The Ohio State University. Columbus, OH. Available at http://www.ling.ohio-state.edu/publications/dissertations

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication, 49*, 28-48.

THANK YOU FOR YOUR ATTENTION

Contact information: *maria-del-mar.cordero-rull@etu.unice.fr*