

# Acoustic and semantic processes during speech segmentation in French

Mar Cordero-Rull<sup>1</sup>, Stéphane Pota<sup>2</sup>, Elsa Spinelli<sup>2</sup>, Fanny Meunier<sup>1</sup>

<sup>1</sup>Université Côte d'Azur, CNRS, BCL, France

<sup>2</sup>Université Grenoble Alpes, CNRS, LPNC, France

<https://doi.org/10.36505/ExLing-2021/12/0014/000487>

## Abstract

We designed two experiments that tested the listeners' perceptual capacities during online segmentation of homophonic word boundaries while processing sentential information. In French, listeners often use variations in fine acoustic indices to detect word beginnings. We measured event-related potentials (ERPs) evoked by phonemically identical sequences, such as *l'affiche* ("the poster") and *la fiche* ("the sheet"), both [lafɛʃ], which were contained in either congruent or incongruent sentences. Results showed that although listeners can detect acoustic variations in homophonic sequences, these may not be salient enough when contextual information is also present. Shifting attention from sentence meaning (Task 1) to lexical information (Task 2), enhanced the listeners' perception of fine-grained acoustic details. Thus, top-down processes are likely to modulate speech perception and segmentation.

Keywords: speech segmentation, sentence processing, acoustic cues, ERP

## Introduction

To achieve successful comprehension, mappings are drawn from acoustic signal into lexical representations in memory (Davis, Rodd, 2011). Word boundaries can be detected by fine-grained acoustic details (Friederici, Wessels, 1993), which will depend on the listener's language. French listeners often use variations of acoustic information in content word beginnings to detect word boundaries. Several studies reveal that they can discriminate homophonic sequences such as *l'affiche-la fiche*, both [lafɛʃ] (Spinelli et al., 2007, 2010). Yet, both acoustic and contextual information are embedded during processing of ambiguous speech (Shoemaker, 2014).

In our study, we explored how French listeners use linguistic cues in processing ambiguous word boundaries within a specific context. To answer this, we presented homophonic sequences (HS) such as *l'affiche-la fiche* embedded in naturally produced sentences. Contextual information was presented under 3 different conditions: control, congruent, and incongruent. In Task 1, we used a semantical judgement task to detect whether French listeners can process fine-grained acoustic details from HS within a given context. In Task 2, we examined the cognitive cost of segmenting ambiguity during sentence processing. In a cross-modal identification task, attention was guided towards lexical instead of sentential information. To assess the listeners'

capacity to discriminate fine-grained acoustic details from HS, the N400 was considered as a relevant neurophysiological marker of interest. In ERP research, N400 is a negative component reflecting difficulties of lexical semantic integration for incongruent compared to congruent conditions, usually found between 200-600 ms after stimulus onset (Kutas, Federmeier, 2011). If fine-grained acoustic information is encoded during speech processing, listeners should be able to detect a semantic violation, yielding an N400 effect for incongruent context. If, however, HS' meaning is accessed through contextual cues, no N400 effect is expected since incongruent context will not be perceived as a semantic violation.

## Method

### Participants

Forty-nine native speakers of French took part in this study (Task 1: 21 participants, 11 females,  $21 \pm 2.42$  years old; Task 2: 28 participants, 22 females,  $22.2 \pm 4.3$  years old). All were right-handed and did not report any psychiatric, neurological, or hearing disorders.

### Stimuli

In both tasks, 144 sentences were presented: 48 control, 48 congruent, 48 incongruent. Control sentences consisted of grammatical sentences containing semantic violations of incoherent target words. For congruent and incongruent sentences, we used 24 pairs of HS in French (e.g., *l'ami-la mie*). They were included in sentences with a semantic constraint that either favored access to the meaning of the presented candidate (congruent) or in favor of the other candidate in the pair (incongruent). Sentences were recorded by a French female speaker, naïve to the study goals. To avoid acoustic differences, homophone candidates were excised from original sentences and introduced in either incongruent contexts (*cross-splicing*) or congruent contexts (*identity-splicing*). Sentence length ranged between 6 and 13 words ( $8.54 \pm 1.48$ ).

### Procedure

In Task 1, a written sentence appeared after a set of 6-13 spoken sentences, after which participants had to decide whether the written sentence was semantically related to the last sentence they have heard. In Task 2, a written word was presented after each auditory sentence and participants had to decide as quickly and as accurately as possible whether the written word was present in the auditory sentence. The selected written words were either present in the sentence (37.1% of the case), absent in the sentence (37.1%), or the target word homophone from the auditory sentence (25.8%).

EEG recordings were acquired using SynAmps2/RT amplifier (Compumedics), recorded from 64 active channels (Neuroscan Quik-Cap Neo Net), following the standard international 10/20 positioning system.

## EEG pre-processing and analysis

EEG data were pre-processed in MATLAB and analysed in R. A 0.1-48 Hz bandpass filter was applied. Epochs were extracted for a span of -200 to 1100 ms relative to article onset. Bad channels were interpolated, and blink and muscle artifacts were corrected. Reference was recomputed to average. Time windows of interest were selected based on global field power (Lehmann & Skrandies, 1980). Selected time-windows corresponded to 0-275, 275-550, 550-870, 870-1100 ms after critical word onset. Using the same time windows across tasks enabled comparisons. Signal amplitude was analysed through permutation tests carried separately for each time window.

### Results in task 1

Between 275-550 ms, a negative activity spread over frontocentral electrodes slightly right lateralized for control context, indicating the presence of an N400 component. It reached statistical significance only compared to congruent context. Compared to incongruent context, differences in the negative amplitude reached significance for central right electrodes (870-1100 ms).

### Results in task 2

An N400 component for control context was observed broadly (550-870 ms); then localized in frontocentral sides (870-1100 ms). For incongruent context, the N400 effect was localized in frontal scalp sides for both 550-870-ms and 870-1100-ms time windows. Significant differences were found between congruent and incongruent contexts over left frontal scalp sides.

## Conclusions and general discussion

In our paradigm, sentential information and HS were pointing to either the same (congruent condition) or different meaning interpretations (incongruent condition). Even though different acoustic indices were associated to HS, no differences were found in the processing of congruent and incongruent contexts (Task 1). Nevertheless, congruent and incongruent contexts yielded an N400 amplitude in Task 2, showing different amplitude sizes relative to their difficulty in integration. Although this effect can also be attributed to an effect of task (Kutas, Federmeier, 2011), results reflect the weighting of multiple segmentation strategies. This reveals the possible influence of a cue hierarchy based on individual saliency during speech processing (Rodd 2018; Mattys, 2003). Top-down processes involving attention may influence the perceptual processing of HS in continued speech. Differences in tasks 1 and 2 suggest that listeners may not use acoustic variations for lexical access when higher-level information (semantic) is salient; whereas fine-grained acoustic details may be used when attention is shifted towards lexical information. Thus, processing acoustic cues may not lead to reinterpretation of meaning when context is

present (Mattys et al., 2005), reflecting an enhancement of top-down contextual information over bottom-up perception.

## Acknowledgements

This work was supported by the French government through the UCA-Jedi project managed by the National Research Agency (ANR-15-IDEX-01) and by the CNRS Service and Research Unit Maison des Sciences de l'Homme et de la Société Sud-Est (MSHS Sud-Est, USR 3566) of the Université Côte d'Azur. We thank Ambre Denis-Noël for her contribution to the project.

## References

- Davis, M.H., Rodd, J. M. 2011. Brain structures underlying lexical processing of speech: Evidence from brain imaging. In Gaskell, G., Zwitserlood, P. (eds.) 2011, *Lexical representation: A multidisciplinary approach*, 197-230. Berlin: Mouton de Gruyter.
- Do Carmo-Blanco, N., et al. 2019. Processing of non-contrastive subphonemic features in French homophonous utterances: An MMN study. *J. Neurolinguistics*, 52.
- Friederici, A. D., Wessels, J. M. I. 1993. Phonotactic knowledge of word boundaries and its use in infant speech perception. *Percept. Psychophys.* 54, 287-295.
- Kutas, M., Federmeier, K. D. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.*, 62, 621–647.
- Lehmann., D., Skrandies, W. 1980. Reference-free identification of components of checkerboard evoked multichannel potentials fields. *Electroencephalogr. Clin Neurophysiol.* 48: 609–21.
- Mattys, S. L. 2003. Stress-based speech segmentation revisited. *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology*, 121–124. Geneva, Switzerland.
- Mattys, S. L., et al. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *J. Exp. Psychol. Gen.*, 134, 477-500.
- Rodd, J. M. 2018. Lexical Ambiguity. In S.-A. Rueschemeyer & M. G. Gaskell (eds.) 2018, *Oxford Handbook of Psycholinguistics*, 120-144. Oxford, UK: Oxford University Press.
- Spinelli, E., et al. 2010. An intonational cue to word segmentation in phonemically identical sequences. *Atten. Percept. Psychophys.* 72, 775-787.
- Spinelli, E., et al. 2007. Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of French elision. *Lang. Cogn. Process.*, 22:6, 828-859.