Exploring the rhythmic segmentation of heard speech using evoked potentials

Annie C. Gilbert[1], Victor J. Boucher[1] & Boutheina Jemel[2]

[1] Laboratoire de Sciences Phonétiques, Université de Montréal, Canada
[2] Laboratoire de Recherche en Neurosciences et Électrophysiologie Cognitive,
Hôpital Rivière-des-Prairies, Canada
`annie.gilbert@umontreal.ca; www.phonetique.info`

Abstract

This study examines, via evoked potentials called closure-positive-shifts (CPSs), how listeners segment heard utterances on-line. The aim was to determine whether marks of rhythm groups in heard utterances can evoke CPSs independent of varying intonation and syntactic structures. Ten subjects were presented with sets of utterances bearing changing intonation and syntax and the results show that CPS is specifically evoked by marks of RGs.

Index Terms: speech segmentation, prosody, speech rhythm, evoked potentials, electroencephalography

1. Introduction

In research on spoken language, the technique of evoked potentials (EPs) offers a central advantage over behavioral measures in observing how heard utterances are segmented on-line. The importance of the segmentation process can be weighed by considering that, to this day, there is no working definition for assumed linguistic units such as "words" (e.g., Dixon & Aikhenvald, 2002[1]). Nor is there any universal cue that supports a division of utterances into units resembling words in writing. As one prominent linguist put it: "words, as is well known are not present in any direct way in the signal" (Jackendoff, 2007, p. 378[2]). The problem is that dominant theories of language are erected on analysis of transcribed speech and assumptions of letter-like phonemes, words, phrases, and sentences, which some critics claim to be conceptually linked to writing[3]. On the other hand, it is clear that utterances are made up of semantically interpretable units. If these do not correspond to entities like words, then it is essential to any theory of language processing to determine what these units are. On this fundamental question, discovered neural components presenting a characteristic CPS provide vital information on the role of prosodic structure in the segmentation process.

The first studies to bear out the link between CPSs and prosody showed that these components are evoked by intonation contours in heard German utterances.[4] These results were interpreted as indicating that CPS reflects a processing of intonational "phrases". Subsequent studies replicated these results with English utterances,[5] but also showed that CPS could be evoked by contours in meaningless series of syllables or hums, that is heard speech without words or phrases. Thus, CPS is not evoked by intonation reflecting syntactic constituents like phrases but by intonational *groups* (IGs) as such, regardless of whether heard series contain any syntax.

On the other hand, the above work did not consider that IGs can contain rhythm groups (RGs). These groups marked by final lengthening and optional pauses can also. RGs are reputed to be universal and this is seen in behaviors such as in the oral recall of digits where all speakers, regardless of language will spontaneously create RGs. However, marks of RGs can be obscured by salient lexical stress in languages like German or English. In fact, in these languages, RGs only become obvious in particular contrastive contexts. For instance, the utterance "the Greek history teacher" would usually reflect a single IG but can be produced with differing RGs, by lengthening "Greek" or –ry of "history" (and adding optional pauses). This timing change creates a detectable difference in rhythm, leading to a different semantic chunking (i.e. lengthening –ry gives "my Greek-history teacher" meaning *my teacher of Greek history*, whereas in lengthening "Greek" one has "my Greek history-teacher" meaning *my teacher of history who is Greek*). But aside from such selective contexts, it becomes difficult to disentangle marks of lexical stress from those marking RGs. This is not the case for languages such as French, which does not have lexical stress. Using this language, then, is particularly useful in determining whether CPSs are evoked by the specific marks of RGs in utterances independently of IGs or assumed syntactic units like phrases or words. With this purpose in mind, the present study explored whether CPSs is robustly evoked by RGs in heard French utterances regardless of the presence of variable IGs and syntactic structures.

2. Method

2.1. Participants

The subjects were ten native speakers of French (6 females, 4 males) aged 19 to 39 years recruited on the campus of the Université de Montréal. All were right-handed individuals with no previous history of neurological or psychiatric disorders. A standard audiometric screening (Pure tone average) of the participants showed that all presented normal hearing at 15 dB HL or better.

2.2. Stimuli

2.2.1. Design of contexts

The target contexts were two sets of 80 utterances where each utterance contained nine monosyllabic lexemes and RGs of 3, 4, and 2 syllables. These two sets present coinciding syntactic structure and IGs, as summarized in Table 1. In Set 1, both the major NP-VP boundary and the IG fall on the end of the second RG, whereas in Set 2, the syntactic and IG boundary fall at the end of the first RG. The prediction was that, regardless of the placement of the major syntactic boundary or IG, CPS would be evoked by RGs.

Table 1. *Summary of contexts used as stimuli*

Set 1	Major synt. bound.:	NP	VP
	Intonation gr. (IGs):		
	Rhythm gr. (RGs):		
Set 2	Major synt. bound.:	NP	VP
	Intonation gr. (IGs):		
	Rhythm gr. (RGs):		
		— Short syll.	— Long syll.

2.2.2. *Stimuli recording*

The stimuli were spoken by a native speaker of Québec French while listening to the continuous playback of a metronome-like digital pacer. The pacer was made of pure-tone beats bearing the intended rhythm and intonation marks for each condition. (Longer beats marking the ends of RGs and F0 resets marking the ends of IGs.). The speaker was instructed to follow the heard patterns as closely as possible. Pre-test observations showed that patterns produced using this entrainment technique contained the desired group-final lengthening for each RG and equivalent IGs to those presented. Several tokens of each utterance were digitally recorded (44,1 kHz, 16 bits) and the experimenter chose the one occurrence that most closely resembled the guides. See Figure 1 for F0 and intensity contours of both sets of stimuli.

2.3. Procedure

The stimuli were delivered via insert earphones (*ER-3A*) and the target contexts were presented along with filler utterances in random blocks of no more than six minutes using *E-Prime*, with rest pauses between blocks. The output intensity of the soundcard was kept constant and peak amplitude of speech was no more than 68dBA at the earphones. Each utterance was also followed by a separate prompt (a monosyllabic lexeme). In the task, the participants were instructed to listen to the utterance with its following prompt and to determine via a key press if the prompted lexeme was present or absent form the utterance. (The results of this task are irrelevant to the present study.) During the test, participants had to fix their gaze on a point displayed on a monitor to minimize eye movements.

2.4. EEG Recording and ERP analysis

The EEG data was recorded through 58 electrically shielded electrodes embedded in an elastic cap (*Easy cap*) according to the enhanced 10-20 system. Two bipolar electrodes placed above and below the dominant eye (vertical EOG) and at the outer canthus of each eye (horizontal EOG) were used to record eye movements and blinks. A left mastoid electrode was used as online reference for all scalp electrodes and AFz was used as ground electrode. The right mastoid was actively recorded as an additional reference channel. The EEG and EOG were recorded continuously with a band-pass from DC to 100 Hz at a sampling rate of 512 Hz, and stored along
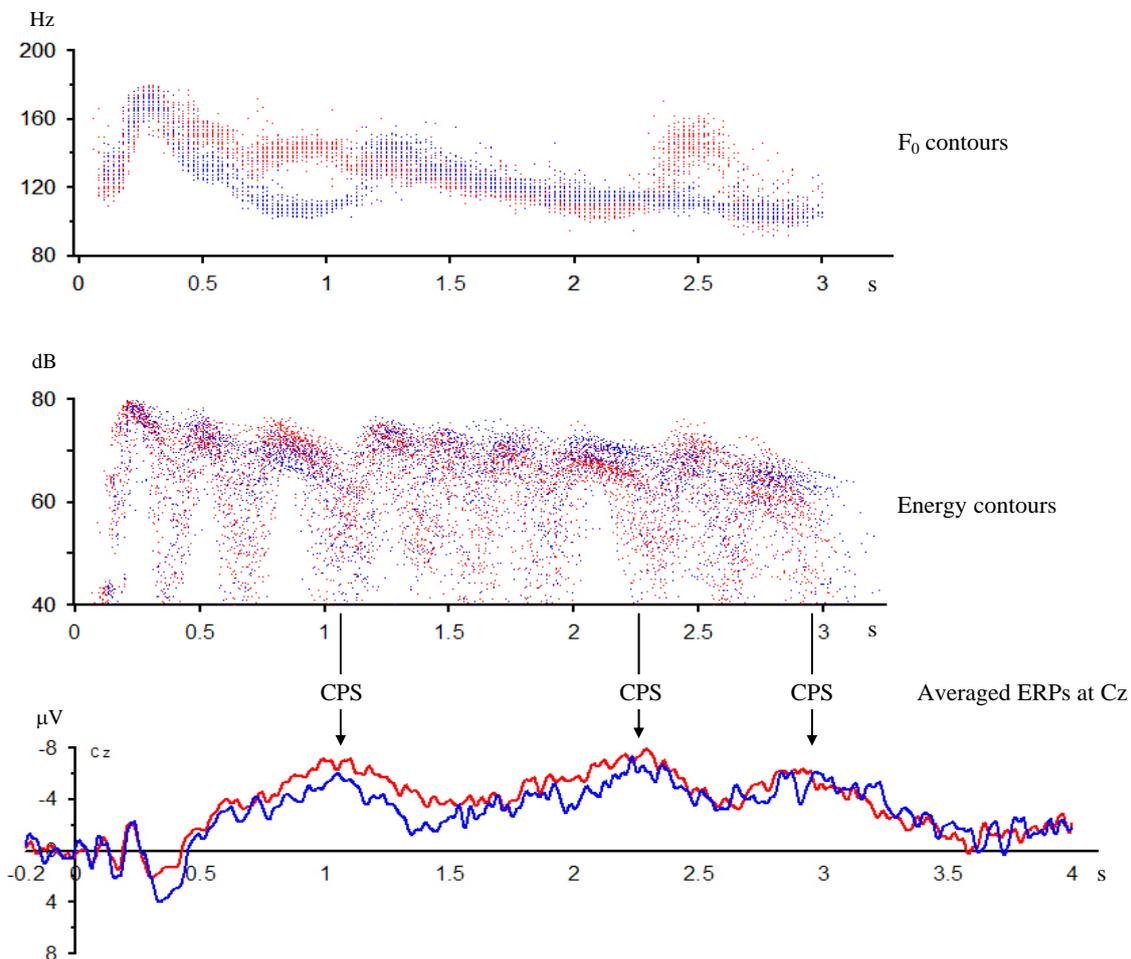

Figure 1: *Stimuli and averaged ERPs at Cz (Red = Set 1, Blue = Set 2)*

with the trigger codes. The EEG signal was filtered using a digital low-pass filter (30Hz) off line and re-referenced to the right mastoid electrode. Segments with artefacts were rejected. Eye-blinks were detected and corrected by subtracting from the EEG the PCA-transformed EOG components for each electrode, weighted according to VEOG propagation factors (computed via linear regression). Artefact-free EEG segments time-locked to the onset of the utterance file were averaged from 200 ms before and 4000 ms after wave file onset for each condition.

3. Results

Averaged ERPs at Cz for the two sets of utterances are presented in Figure 1 along F0 and energy contours from every stimuli. This illustration allows a direct comparison of the stimuli from both sets of contexts containing variably positioned boundaries of IG and syntactic constituents, but sharing the same rhythmic structure

As can be seen in Figure 1, the main finding is that, regardless of the boundaries of IG or syntactic constituents, CPSs are specifically evoked by RGs.

4. Discussion and conclusion

The above exploratory findings accord with previous observations showing that CPSs can be evoked independently of intonation contours and syntax. In the present case, it is the RGs and not the major syntactic or tonal breaks that create a response. Moreover, it should be seen that the results offer an indication of the on-line segmentation of speech flow and that this segmentation appears to bear on rhythm units.

5. References

[1]   Dixon, R.M.W. and A.Y. Aikhenvald, eds. *Word: a cross-linguistic typology*. 2002, Cambridge University Press: Cambridge.

[2]   Jackendoff, R., *Linguistics in cognitive science: the state of the art* The Linguistic Review, 2007. **24**: p. 347-401.

[3]   Coulmas, F., *Units of speech and units of writing*, in *The Writing Systems of the World*. 1989, Basil Blackwell: Oxford, UK/Cambridge, Massachusetts. p. 37-54.

[4]   Steinhauer, K., K. Alter, and A.D. Friederici, *Brain potentials indicate immediate use of prosodic cues in natural speech processing*. Nature Neuroscience, 1999. **2**: p. 191-196.

[5]   Pannekamp, A., et al., *Prosody-driven sentence processing: An event-related brain potential study.* Journal of Cognitive Neuroscience, 2005. **17**: p. 407-421.