Size of rhythm-groups affects the memory trace of heard words in utterances: Results from a pilot study using evoked potentials

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

[1]Laboratoire de Sciences Phonétiques, Département de linguistique et de traduction
Université de Montréal, Québec, Canada

[2]Laboratoire de Recherche en Neurosciences et Électrophysiologie Cognitive
Département de Psychiatrie, Université de Montréal, Québec, Canada

annie.gilbert@umontreal.ca   www.ling.umontreal.ca/phonetique

Abstract

Since speech involves fleeting acoustic events some portions of heard utterances must be parsed and stored on-line in order to be interpreted. Our previous work has shown that in recalling novel series, rhythmic grouping of up to 4 syllables facilitates memory, and this corresponds to size limits on rhythm groups observed in speech. Such results suggest that prosodic groups can correspond to the natural parsing frameworks of memory processing. The present study further investigates this correspondence by means of an event-related potential technique. It has been established that the quality of memory traces of words is associated with variations in the amplitude of the N400 component. We show that variations in the amplitude of the N400 confirm for one subject that words presented in rhythm groups that do not exceed a 4-syllable limit present a better quality of memory trace than words presented in groups of 5 syllables, which rarely occur in speech and do not facilitate recall. Though the present observations are preliminary, our results converge on the point that prosodic grouping can be linked to mnemonic processes and that restrictions on memory may underlie size limits on prosodic groups.

1. Introduction

It is generally recognized that prosodic structure is not entirely predictable from syntax and that independent "size-effects" appear, suggesting that prosody operates by some independent principle ([1-4]). To illustrate size effects, consider the following sentences used by Jusczyk[5] to illustrate changes in the boundaries of prosodic groups (which the author indicated by "/"):

    (1)   He ate/the cake
    (2)   Thomas/ate the cake

Some fluctuation in grouping is likely in an utterance like (2) (e.g. "Thomas ate/the cake" seems just as likely). But where the grouping becomes highly predictable is when one lengthens constituents as in (3)

    (3)   The neighbour's boy/ate the cake

A prosodic boundary after "ate" in this case would be unusual. Thus, in increasing the size (number of syllables) of constituents, prosodic grouping becomes predictable, and this is basically what we found in an earlier report on French rhythm groups[3]. The point is, such effects are not predictable by reference to major or minor syntactic divisions and tends to question the dominant assumption that prosodic structure reflects or derives from syntax. Moreover, while some authors have proposed formal devices and complex algorithms that basically re-align "prosodic constituents" to assumed (surface) syntactic forms[6; 7], Gee and Grosjean[8] note, that formal models and algorithms "are not in themselves interesting unless they are steps on the way to a model of the human mind" [p.449]. In other words, algorithms or a priori formal devices do not as such explain the nature of prosodic structure and size effects.

Considering the above examples, one should note that grouping within such short sentences may not correspond to a pitch reset marking an intonational group (IG) but to a temporal changes occurring at the ends of what can be termed rhythm groups (RGs). In our observations of these RGs, we have preferred to use French speech rather than a language with lexical stress such as English. The reason is that lexical stress presents a central confound in isolating temporal RGs. For example, in the above utterance (1), lexical stress can overlap a group-final mark. In French, however, there is no word stress and this presents an advantage in observing the lengthening marks associated to RGs, which are not coded in language and which can occur internal to an IG. What is particularly revealing is that size-effects on RGs can present a general structuring principle, and our previous work has shown that there is a correspondence with respect to grouping effects on the recall of speech material.

A number of reports point toward the existence of constraints on the number of syllables of prosodic groups, which may generalize across languages. Studies of several languages have shown that RGs and inter-stress intervals tend to have an average upper-limit of about 4 syllables[7; 9-11] whereas IGs have an average upper limit of 8 syllables (based on French speech[12]). Inasmuch as RGs have one stress mark, the tendency to limit inter-stress intervals would, in principle, reflect a general tendency to limit RGs. What is particularly revealing is that this upper-limit matches a restriction on grouping effects on serial recall. That is, it has been known for some time that the recall of lists of words like digits or letters is facilitated when items are presented in groups of at most four items. This effect has been shown to be applicable to heard lists presented with various grouping marks[13-17]. Similar size-limits on memory processing of series have been suggested by Cowan[18] who found that the focus of attention in humans is restricted to about 4 items. A direct link between this size limit to grouping effects and prosodic RGs in speech was demonstrated by Boucher[3] (see also[19]). In particular, it was shown that serial recall of novel series of syllables is facilitated by RG of at most 4 syllables, and this corresponded to a restriction on RGs observed in a task involving semi-spontaneous speech[3].

At first glance the aforementioned finding is difficult to interpret from a linguistic perspective, since prominent authors like Chomsky[20] explicitly excluded memory from having any role in understanding language competence. But one can reason that learning a language requires the ability to reproduce heard patterns and series of prosodically structured sounds constituting new words or expressions. In this light, constraints on immediate memory likely reduce the number of grouping patterns that would be learned by a speaker. Thus, the link between constraints on the size of grouping patterns and memory (or associated processes like attention) appears logical and straightforward: long groupings that are difficult to reproduce and that do not facilitate memory of speech will not tend to be learned and will be less likely to be used in parsing utterances.

To further explore the link between memory and prosodic structures, an event-related potential (ERP) task was designed with the purpose of observing the possible effects of the length of RGs on the quality of the memory trace of words. The task was elaborated by considering our earlier findings that RGs in speech do not tend to exceed 4 syllables, and that similar limits apply to groups that facilitate immediate memory of heard series. On this basis, it was hypothesized that words within RGs matching the 4-syllable limit would present a better memory trace than words within larger RGs. Using the ERP technique, the "quality" of the memory trace can be evaluated through the amplitude of the N400 triggered by the active retrieval of information from presented stimuli. Previous studies have established that the quality of a memory trace is inversely related to the amplitude of N400: comparatively smaller amplitudes occur for stimuli with a better memory trace. (For the link between N400 effect and memory see, among others [21-23]) In the present pilot study, French sentences were used to avoid the effect of lexical marks. In the task subjects are required to say if a word is present in a previously heard utterance. The main prediction was that N400 amplitudes would be smaller for words within 4-syllable RGs than for words within 5-syllable RGs, which exceed the general size limit on groups in speech.

2. Method

2.1. Subject

The present pre-test was performed with one right-handed female speaker of Québec French (26 years of age). The subject had no history of hearing or speech related problems and was under no medication.

2.2. Stimuli

2.2.1. Sentence contexts

Two sets of 26 French sentences were used. These sentences contained RGs of 4, 5, and 2 syllables for a total of 11 syllables (with the 2-syll. RG always last). The two sets differed in the order of the 4- and 5-syllable groups (one set contained groups 4:5:2, the other 5:4:2). The syntactic constituents of the sentences presented similar distributions of grammatical categories, as specified in Table 1, and all words in the sentences were mono-syllables. As for the target words used as prompts for "yes"/"no" responses (see *Procedure*), half were taken from the first RG and half from the second. Seventy-two filler/distractor sentences were included and these sentences contained varying RGs. Target words were

also taken from these contexts. In all, then 124 sentences were used.

Table 1: *Grammatical categories as function of RG length.*

RG length	Syllable order				
	1st	2nd	3rd	4th	5th
5 syll	Det. or Prep.	Det. or Adj.	Adj.	Noun	Adj.
4 syll	Det. or Prep.	Det. or Adj.	Noun	Adj.	
2 syll	Verb	Adj. or Adv			

2.2.2. Elaboration of acoustic stimuli

The stimuli were constructed by asking a male, native speaker of Québec French to read and then say the aforementioned sentences while listening to a metronome-like (digital) signal providing guides for rhythm and intonation. These guides were sequences of pure-tone beats with longer beats marking the ends of RGs, and with falling pitch and resets marking IGs. The speaker was instructed to follow the heard patterns as closely as possible. Using headphones, the speaker listened to a continuous playback of the metronome while he repeated each sentence. The guides were build in order elicit the production of RGs with constant durations (4-syll RG = 1 150 ms, 5-syll. RG = 1 400 ms, 2-syll. RG = 650 ms.) marked by the lengthening of the last syllable making it 1.6 times longer than non-final syllables[24]. A 50 ms silence was added between RGs to ensure their clear separation, resulting in sentences of a fixed duration of 3 300 ms. Pre-test observations showed that patterns produced using this entrainment technique contained the desired group-final lengthenings for each RG and equivalent IGs to those presented. The tonal contour was controlled to make sure that no pitch movement or reset occurred between the first and second RGs to avoid the presence of a tonal group (a reset occurred only between the second and last RG). The speaker produced several tokens of each sentence and the experimenter chose the one occurrence that most closely resembled the guides. The target words were obtained by splicing units from the recorded sentences. <

2.2.3. Procedure

The stimuli were presented via insert earphones (*ER-3A*; Etymotic Research, with *E.A.Rlink* foam eartips) and output was calibrated to reach maximum amplitude of 74 dBA at the subject's ear. Presentation was randomized via a function of the presentation software (*E-Prime*, Psychology Software Tools) and every target sentence was presented twice along with filler sentences. In the test, the subject was asked to pay attention to the sentences followed by a target word and decide, as fast and as accurately as possible, whether or not the word was present in the heard sentence by pressing on left-hand or right-hand keys of a response-box. However, in this study, response times and accuracy were not recorded.

2.3. ERP data collection and analysis

The ERP 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 earlobe electrode was used as a reference for all scalp electrodes and (AFz) was used as ground electrode. The right earlobe 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 1024 Hz, and stored along with the trigger codes. The EEG signal was filtered using a digital band-pass filter (0.03-30Hz, 1501 pts window) off line and re-referenced to the right earlobe electrode. Eye blinks being rare in the recorded data, no correction was applied to remove them from the recordings. Electrodes T7, TP7 and FPZ had to be removed from the analyses because of major DC drifts. EEG segments time-locked to the onset of the target-word onset were averaged from 200 ms before and 1000 ms after target word onset for each condition.

3. Results

Average ERPs at Fz, Cz and Pz for target words are presented in Figure 1 as a function of their RG of origin (first or second RG). This allows the comparison between speech-size and non speech-size RGs in both sentence initial and internal positions. The Fz electrode site is reported since the frontal region has been shown to be associated with attention and working memory. On the other hand, Cz and Pz were selected because N400 have also been shown to occur at these sites (see, among others,[21; 23]). The ERP signal was smoothed using a 50-pt running average to clarify the presentation of effects. It is important to note that the graphs were adjusted in their time-reference to better reflect the actual *P-center* (perceptual-center)[25; 26] of the mono-syllabic words. Thus, time "0" in Figure 1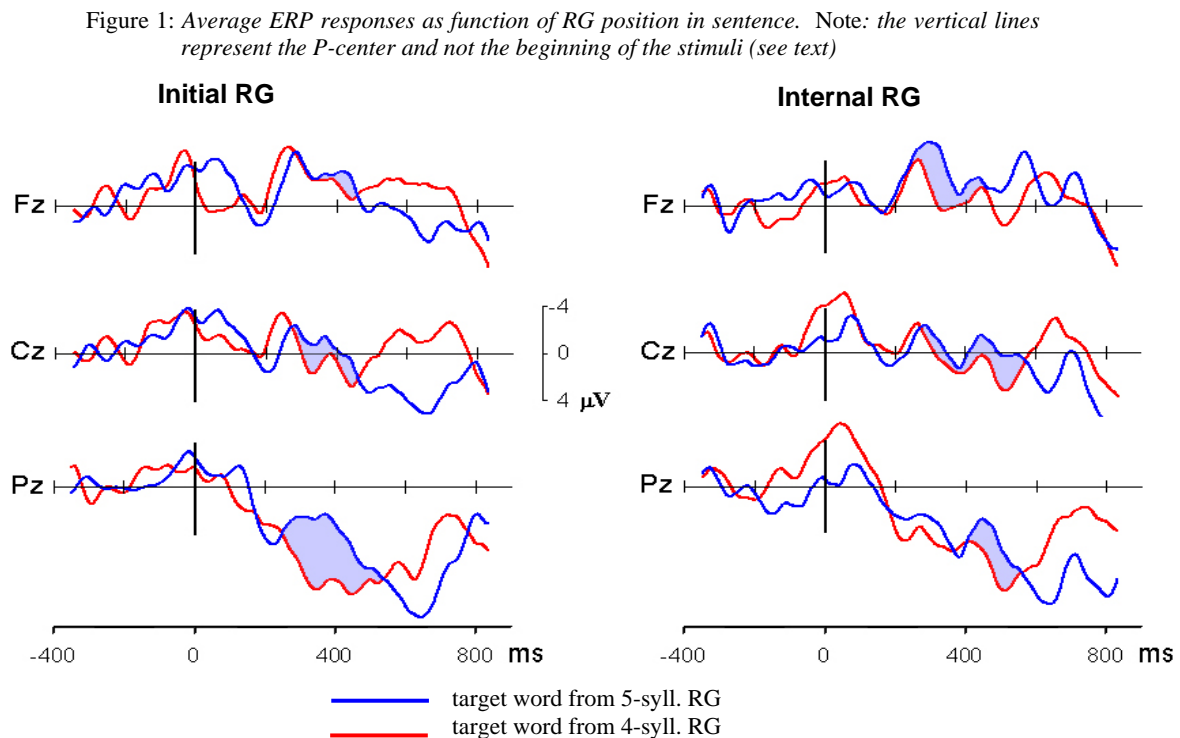 represent the moment of a word's P-center and not the onset of the stimuli. This provided a means of controlling for possible confounding effects of differences in the absolute duration of target-words.

As can be seen in Figure 1, the main finding is that N400 amplitudes are always greater for words from 5-syllable RGs than for words from 4-syllable groups, and this applies regardless of group position. This offers some preliminary confirmation of our prediction that the memory trace for words is superior in quality when the words are in a RG that reflects the size of groups occurring in speech, which also corresponds the size of groups that benefit list recall. It should be noted that the difference between N400 amplitudes is also larger at posterior sites, which agrees with an effect reported by Finnigan et al.[23] for visually presented words.

4. Discussion and Conclusion

The above observations offer initial support for the view that, when listening to utterances, the mnemonic trace of words is superior when the units occur within rhythm groups that do not exceed a limit-number of syllables. When groups exceed 4 syllables, the memory trace of words decreases in quality. The quality of the memory trace can be related to the amplitude of the N400, as has been shown by Finnigan et al.[23] In that study, subjects were first asked to pay attention to lists of words presented on a computer screen. In the list, some words were presented three times while others were presented only once. The words presented three times strengthened the memory trace and recall, and this was associated with a lower-amplitude N400 compared to amplitudes for words presented once only.

While the above results are preliminary, they point to the possible role of groupings of given size in on-line processing of utterances. Acoustic stimuli are fleeting events so some portion of utterances must be parsed and stored on-line. What is revealing in the present study is that RGs in speech that appear to contribute to the mnemonic trace of words are of the same size as groups that facilitate list recall. When the groups

Figure 1: *Average ERP responses as function of RG position in sentence.* Note*: the vertical lines represent the P-center and not the beginning of the stimuli (see text)*

exceed a given size limit of 4 syllables, the mnemonic trace weakens, and similarly, it has been shown that the recall of lists is facilitated by groupings of no more than 4 items. Pursuing the investigation of this correspondence may reveal that the underlying principle of prosodic structures is related to constraints on memory processing.

5. References

[1] Hirst, D. (1993). Detaching intonational phrases from syntactic structure. *Linguistic Inquiry, 24*(4), 781-788.

[2] Scholes, R. H. (1971). *Acoustic cues for constituent structure*. The Hague, The Netherlands.: Mouton & Co.

[3] Boucher, V. J. (2006). On the function of stress rhythms in speech: Evidence of a link with grouping effects on serial memory. *Language and Speech, 49*, 495-519.

[4] Gilbert, A. C., & Boucher, V. J. (2006). Syntax and syllable count as predictors of French tonal groups: Drawing links to memory for prosody. In *Proceedings of the 3rd International Conference on Speech Prosody* (pp. 289-292). Dresden, Allemagne.

[5] Jusczyk, P. W. (2001). Bootstrapping from the signal: some further directions. In J. Weissenborn & B. Hohle (Eds.), *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition* (pp. 3-23). Amsterdam: John Benjamins Publishing.

[6] Selkirk, E. O. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody, theory and experiment: Studies presented to Gösta Bruce* (pp. 231-262). Dordrecht, The Netherlands: Kluwer Academic.

[7] Monnin, P., & Grosjean, F. (1993). Les structures de performance en Français: Caractéristiques et prédiction. *L'Année psychologique, 93*, 9-30.

[8] Gee, J. P., & Grosjean, F. (1983). Performance Structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*, 411-458.

[9] Dauer, R. M. (1983). Stress-timing and syllabe-timing reanalyzed. *Journal of Phonetics, 11*, 51-62.

[10] Martin, P. (1999). Intonation of spontaneous speech in French. *Proceedings of the 14th International Congress of Phonetic Sciences*, 17-20.

[11] Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics, 19*, 193-212.

[12] Gilbert, A. C., & Boucher, V. J. (2005). *Size-constraints on intonation groups in speech: Evidence of an independent syllable-count principle*. Paper presented at the 149th Meeting of the Acoustical Society of America, Vancouver, Canada.

[13] Maybery, M. T., Permentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: implications for models of serial verbal memory. *Journal of Memory and Language, 47*(3), 360-385.

[14] Reeves, C., Schauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1638-1654.

[15] Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *The Quarterly Journal of Experimental Psychology, 49A*(1), 116-139.

[16] Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology, 9*(special issue), S5-S22.

[17] Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(3), 469-479.

[18] Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science, 15*, 634-640.

[19] Boucher, V. J., & Gilbert, A. C. (2005). Constraints on immediate memory for prosody: Evidence of a correspondence with the prosodic organization of speech. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* (pp. 328-330). Mahwah, NJ: Erlbaum.

[20] Chomsky, N. (1967). The formal nature of language. In E. H. Lenneberg (Ed.), *Biological foundations of language* (pp. 397-442). New York: John Wiley & Sons, Inc.

[21] Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30*, 47-61.

[22] Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

[23] Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP "old/new" effects: Memory strength and decisional factor(s). *Neuropsychologia, 40*, 2288-2304.

[24] Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics, 4*, 183-198.

[25] Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics, 30*, 247-256.

[26] Fowler, C. A., Whalen, D. H., & Cooper, A. M. (1988). Perceived timing is produced timing: A reply to Howell. *Perception & Psychophysics, 43*, 94-98.

6. Acknowledgments