

Investigations of speech segmentation: addressing the writing bias in language research

Victor J. Boucher and Annie C. Gilbert
Laboratoire de sciences phonétiques, Université de Montréal, Canada
<https://doi.org/10.36505/ExLing-2008/02/0012/000071>

Abstract

We address the problem of defining universal processes of speech segmentation in view of criticisms that conceptual linguistic units derive from western writing. A synthesis of our recent experimental studies is presented bearing on processes of serial-order and rhythmic grouping. First, on how serial-order operates, we use EMG and speech-motion data to show that “consonant-vowel” orders reflect contraction-relaxation cycles (not separate segments). Second, on how sequences of sounds come to form units in language learning, we discuss behavioural data suggesting a link between rhythm groups in speech and grouping effects on memory of speech sounds. Preliminary EEG data is then presented to substantiate the view of an on-line parsing of rhythm groups with effects on memory traces of lexemes.

Key words: speech, segmentation, universals, physiology, history of linguistics

The writing bias and the need to define segmentation

The importance of defining processes of segmentation can be weighed by considering the weak validity of formal linguistic units. Historically, though principles like “phonological distinctiveness” played an essential epistemological role in defining features, one finds no principle forcing the view that features occur in “bundles” as in letters of IPA, or that meaningful elements (morphemes) occur in “word” units as in text. In fact, several critics have warned that concepts of *phoneme* and *word* implicitly refer to European writing and orthographic codes (e.g. Coulmas, 1989; Linell, 2005). But while critics point to the fundamental problems arising from a writing bias in conceptualizing units, little research is devoted to defining the processes by which features are ordered and learned in multisyllabic groups. On these issues, we provide a synthesis of our recent work on segmentation, which bears centrally on the processes underlying serial order and rhythmic grouping.

Defining the process of serial-ordering

Using IPA to represent speech can foster a conception that serial order follows letter-like “segments”. Such conception pervades not only linguistic theory but also neural network and speech-motor models. All assume that, in producing a syllable like /pa/ neural influx to articulators, as can be observed by EMG, is organized serially with closer muscles being activated before

openers in line with notions of consecutive consonant and vowel segments. However, Boucher (2008) showed that consecutive activation may not occur. Figure 1 from that study illustrates the activity of the main muscles involved in aperture motions of the lips and jaw. Note that, in creating close-open cycles of the lips while clenching the teeth, activity appears for lip-openers (depressor labii m.). Conversely, in creating close-open cycles of the jaw with lips sealed, activity appears for jaw openers (the anterior digastric m.). However, in creating close-open cycles in speaking /papapa/, activity of openers is not present. This questions the notion of separate activation by reference to [+cons] and [-cons] phonemes and suggests instead a serial ordering where passive factors such as tissue elasticity affects openings.

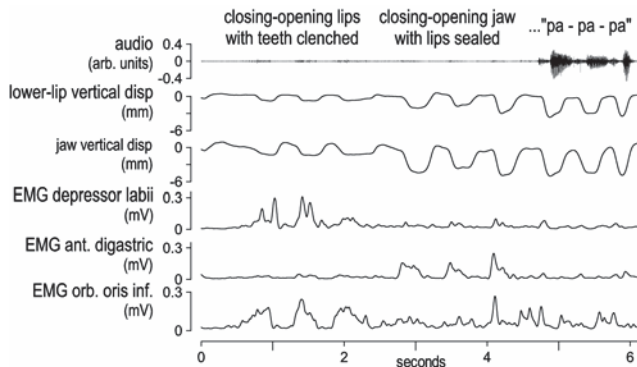


Figure 1. EMG for close-open cycles of the jaw and lips during non-speech and speech. Adapted from Boucher, 2008.

This latter suggestion was investigated by hypothesizing that, if opening motions reflect passive spring-like behavior, then Hooke's law would apply: force applied in compressing a spring leads to an opposite elastic force, which can move a mass at a speed and distance proportional to system constants. We took oral pressure and bilabial compression during closing as indices of applied force and predicted that velocity and distance of lip and jaw opening would be linearly related to applied force. Strain gages were used to monitor lip and jaw motion and 3 subjects were asked to produce series *papa...* and *baba...* with increasing loudness so as to emphasize changes in lip compression and oral pressure. Figure 2 presents an example of the results: both force indices showed such linear relations with speed and range of opening for *ba* and *pa* series, as predicted. In short, close-open cycles of motion in articulators may not reflect serial activation for assumed [+cons] and [-cons] phonemes as currently assumed in motor theories. Instead serial control follows cycles of contraction and relaxation with opening motion being largely driven by passive factors. One should also note that this view of serial-ordering conforms to the perception of speakers who do not know alphabet writing. These speakers can easily count "syllable cycles" but not letter-like phonemes (see Boucher, 1994 in Boucher, 2008).

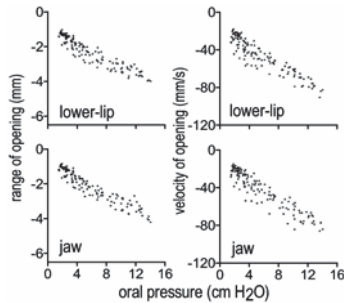


Figure 2. Example of linear relationships between an index of closing force (i-o pressure) and both the speed and range of opening motion, suggesting elasticity effects. Adapted from Boucher, 2008.

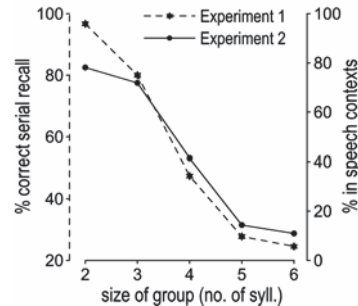


Figure 3. Effects of size of groups on recall (*Experiment 1*) and frequency of occurrence of group sizes in speech contexts (*Experiment 2*). Adapted from Boucher, 2006.

Defining the grouping process in learning language forms

The above defines a basic process of serial ordering. However, specific sequences of speech sounds come to consolidate as forms in language learning. This implies a capacity to parse and hold in memory a given chunk of speech. What is the extent of such parsing and are there any marks of this process in speech? One clue appears when speakers recall novel series of syllables or digits. In such tasks, groupings arise, and it is known that recall is enhanced when lists of items are presented in groups of 3 or 4. What is intriguing is that there is also a tendency in speech to create rhythms that do not exceed 3 or 4 cycles (Boucher, 2006). Considering this coinciding “size-effect,” we explored the link between grouping effects on recall and rhythm groups (RGs) in speech by two experiments involving 40 French speakers.

In *Experiment 1* subjects had to recall heard series of 7 non-sense syllables. Two sets of stimuli were constructed from monotone speech. The first presented set were arrhythmic series of unstressed syllables (*s*); the second set were series with long syllables (*S*), one placed at the end of a sequence and one placed at varying internal positions so as to create rhythms of 2 to 6 syllables (i.e. *s S s s s s S*, *s s S s s s S*, *s s s S s s S*, etc.). In *Experiment 2*, sentences with subject and verb-complement phrases of 2 to 6 syllables were visually presented. Subjects had to say a context and then repeat it twice from memory, once normally and once using glottal stops to mark rhythm. For instance, we used sentences with short and long compound names (*Pierre-Paul part mercredi matin*; *Marie-Antoinette part mercredi matin*) to determine how rhythmic grouping operates when constituents vary in length. The results of *Experiment 1* showed that only groups not exceeding 4 syllables significantly enhance memory compared to arrhythmic

series. More importantly, as seen in Figure 3, the decreasing benefit of long groups in recall appeared to correlate narrowly with the decreasing frequency of long RGs observed in the speech contexts of *Experiment 2*.

While the above results support the hypothesis of a link between groupings that benefit memory and RGs in speech, they do not directly demonstrate an on-line parsing of the groups in listening to utterances. On this problem, a preliminary study using the EEG technique of evoked potentials was conducted (Gilbert et al, 2008). Controlled contexts were presented to a listener consisting of 50 utterances bearing RGs of 4 and 5 syllables *within* (and not at the boundaries of) intonation groups, and these were presented along with distractor utterances. The task consisted of responding to the presence/absence of a given lexeme in the utterance. The results revealed known "closure positive shifts" evoked by length marks of RGs while the subject was listening to the utterances. Also, decreases in N400 were found for target lexemes previously presented in the short RGs, suggesting that, compared to 5-syllable RGs, shorter groups benefit memory traces of lexemes. These results confirm an on-line parsing of heard utterances by reference to RGs and suggest that group-size affects memory of forms within the groups.

Prospective conclusion

It should be noted that the above processes of segmentation are language-independent. They can give rise to forms which may be variably termed and have varying roles in different language systems (e.g. syllables, mora, lexemes, concatenations, etc.). However, applying language-relative terminology and analyzing transcriptions that assume letter-like segments and word divisions may not capture the universal process at the source of the forms.

References

- 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, 495-520.
- Boucher, V. J. 2008. Intrinsic factors of cyclical motion in speech articulators. *Journal of Phonetics*, 36, 295-307.
- Gilbert, A. C., Boucher, V. J. and Jemel, B. 2008. Size of rhythm-groups affects the memory trace of heard words in utterances. In P. A. Barbosa, S. Madureira and C. Reis (Eds.), *Proc. of the Speech Prosody 2008*, 379-382. Campinas, Brazil.