SERIAL-ORDER CONTROL AND GROUPING IN SPEECH: FINDINGS FOR A FRAME/CONTENT THEORY

Victor J. Boucher

Laboratoire de sciences phonétiques, Université de Montréal
victor.boucher@umontral.ca

ABSTRACT

Frame/content (F/C) theory [1] offers a working rationale of the rise of serial-order control and forms without assuming *a priori* units. A synthesis of our recent work is presented with the purpose of refining this rationale on two points. First, observations of contraction activity and passive elasticity suggest that basic frames of serial-order control correspond to contraction-relaxation cycles not present in non-speech motions such as mastication. Second, on explaining prosodic grouping, results show a relationship between "size effects" on such patterns and grouping effects on recall. Converging evidence suggests that grouping may arise from capacity limits on attention processes of short-term memory.

Keywords: serial-order, rhythm, memory

1. INTRODUCTION

In the history of phonetics, the linguistic criterion of "phonological distinctiveness" played an essential epistemological role in restricting the object of speech observations. At one time, this object was limited to uncovering the physical correlates of *distinctive features* and, since then, linguistic concepts have profoundly influenced phonetics. This influence includes the general view that features are ordered in phoneme segments and grouped into (phonological) words. Yet, there is no epistemological principle that forces such a conceptualization of speech. In fact, historians of linguistics have warned that definitions of phonemes and words implicitly refer to writing [2,3]. This has had central consequences in linking phonetic observations to formal constructs. In research on language development especially, it is often assumed that infants *must* uncover phonemes and words before they develop syntax, even though input utterances offer unreliable support for such units (hence the "bootstrapping" problem). Also troubling are reports that certain cultures or illiterate speakers who do not know alphabet writing have no concept of phonemes or words ([2,3]). Thus, while phonetics devotes much attention to features and their motor-sensory basis, questions of how serial ordering operates and why prosodic groups fail to match assumed units remain largely unaddressed. One exception is F/C theory [1], which offers a rationale of both the rise of serial-order control and linguistic forms. On these two fundamental points, the present paper offers a synthesis of our recent findings aimed at refining the F/C rationale.

2. SERIAL ACTIVATION IN SPEECH

F/C theory does not assume that speech motions are sequenced in terms of phoneme strings. It suggests instead that serial-order originates from cyclical motions of mastication. As MacNeilage [1] notes, the ability to control oral cycles of closing and opening is consistent with the growth of specialized functions in the human brain. However, a slightly different perspective on the origin of the cycles can be inferred from observations of serial activity in articulators.

Using electromyography (EMG), we explored [4] the serial activation of the main muscles involved in aperture motions of the jaw and lips including the *anterior digastric* m. (AD), the *depressor labii* m. (DL), and the *orbicularis oris* inf. m. (OOI). Fig. 1 provides an example of what was observed. For non-speech close-open cycles of the lips (with clenched teeth) or the jaw (with lips sealed), activity in the abductors (AD and DL) is clearly present, but not during speech production of *papa…* (to the right of Fig. 1). Thus patterns of activation in speaking do not reflect those of chewing-like motions, but show instead a control specific to the production of sounds. Nor do the patterns support a segment-by-segment control where a feature, say [+cons], activates a closing gesture, and [-cons] then activates a release motion. In fact, there is no influx to muscles for the opening movements so that one is led to view that openings may involve intrinsic factors of elasticity.

relaxing tension at pressure rises presents an intrinsic timing mechanism of serial activation.

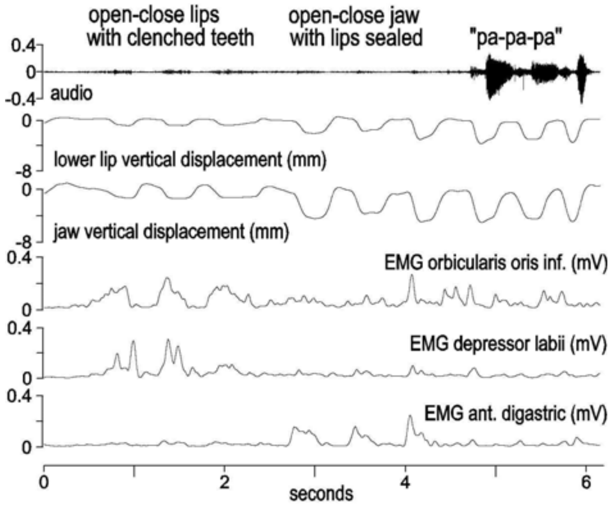

Figure 1: Lip and jaw motions with underlying (rectified, smoothed) EMG activity of abductors and adductors.

On this basis, it was reasoned that, if jaw and lip opening reflect passive spring-like properties of tissues, then certain principles would apply. By Hooke's law, force applied in compressing a spring leads to an opposite elastic force, which can move a mass at a speed and distance that is proportional to system constants. By reference to this principle, we took oral pressure and bilabial compression during closing as indices of applied force. It was predicted that, if passive elasticity contributes to opening motions, then velocity and range of opening would be linearly related to applied force.

Measures were performed via strain gages on the lower lip and jaw. Three subjects were asked to produce series *papa…* and *baba…* with increasing loudness so as to emphasize changes in lip compression and oral pressure. An example of the results is given in Fig. 2. Overall, the data showed that indices of applied force presented strong linear relations with speed and range of opening, confirming that intrinsic elasticity can underlie release kinematics.

Observations by Ostry & Munhall [5] suggest that elastic force may also operate on opening movements of the tongue, and such interpretation is compatible with repeated observations of the co-contractions of "consonant" and "vowel" at the onset of close-open cycles (see [3,4]). One central implication is that, contrary to mastication, tissue elasticity can be involved in producing opening motions in speech. Thus, serial-order control in speech may originate from cycles of contraction and relaxation. Humans are the only primates to control oral pressure in producing sounds [6] and

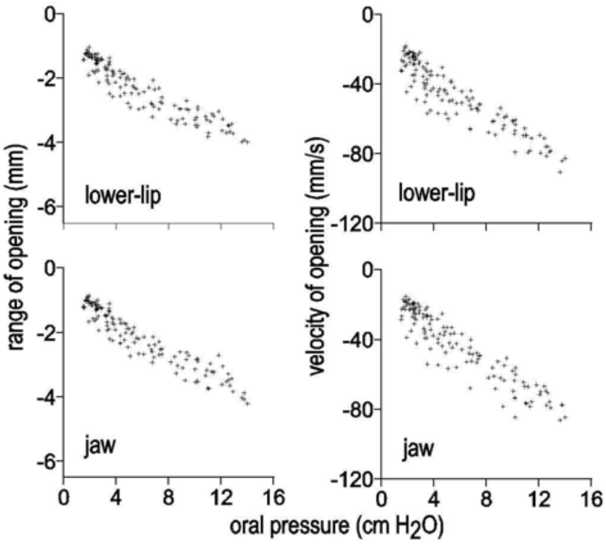

Figure 2: Example of linear relations observed between indices of closing force (here oral-pressure) and kimenatics of opening in speech at varying loudness. Such relations suggest spring-like behavior [4].

3. GROUPING AND "SIZE EFFECTS"

The above mechanism of serial control does not as such explain the second problem of why series of motions are organized in prosodic *groups.* Within F/C theory, reduplicative babbling (at 6-8 months) provides the framework of serial-order control allowing the learning of heard "words". However, the theory does not consider the role of grouping in this learning process, nor the problem of linking speech to assumed word and syntactic units.

On this latter issue, numerous studies have attempted to show that the statistical distribution of marks like lexical stress shapes memory and leads to a conception of units like words (e.g., [7]). Yet, not all languages have "word stress" and some do not divide words [2]. There are also frequent mismatches between prosodic groups and assumed syntactic units. For instance, well known "size effects" occur where rhythm and intonation groups extend across major syntactic divisions (such as subject and verb) when constituents involve small numbers of syllables. It is not clear that such cases are less likely to bias memory than "correct" marks of assumed units like words. However, some authors suggest that prosodic groups of a given size may enhance memory of morpho-syntactic regularities and associations (see in [7]). In fact, it is well attested in memory research that presenting lists of items in groups of three or four benefits serial recall-–though no study has examined whether such effects extrapolate to *prosodic* groups in learning series of speech sounds.

The question thus arises of whether size effects on prosodic structures above the syllable correspond to groupings that facilitate short-term memory (STM). We investigated this link [8] using French speech since, in this language, rhythm groups bear a final lengthening that *is not lexically coded*. This avoids confounding grouping marks with lexical stress. Two experiments were devised involving 40 native speakers.

3.1. Revealing size effects on rhythm groups

We first examined the production of grouping marks in given sentences where the length of subject and verb-complement was made to vary in numbers of syllables. Each phrase contained two nouns, and the contexts were controlled for symmetry effects. For example, subject phrases bore compound names of two syllables (*Pierre-Paul part mercredi matin*) to five syllables (*Marie-Antoinette part mercredi matin*), and similar size changes were applied to verb-complement phrases. In the test, the speakers had to read a displayed sentence, and then say it twice from memory, once normally, and then by humming syllables with glottal stops. With this echoing procedure, glottal beats conformed to the timing of syllables in normally uttered sentences and this facilitated the identification of groups.

Acoustic analyses showed that, within noun phrases containing two or three syllables, internal lengthening marks appeared in 20% of the cases (depending in part on whether or not the verb possessed a mark). But with phrases exceeding four syllables, internal marks occurred in 70 % of the trials. This fragmentation did not reflect intonation groups. For instance, French speakers would not normally divide a name like *Marie-Antoinette* by resetting F0, though marks of rhythmic grouping appeared in such cases and not in shorter names like *Marie-Pierre*. Also, the finding of a limit of about four syllables conforms to Martin's [9] statistics on spontaneous French speech, where it was found that groups do not tend to exceed of 3.6 syllables on average.

3.2. Size limits in grouping effects on STM

The second experiment examined how rhythms of varying size affect STM of novel series. These series were seven CV syllables assembled from monotone productions. One set contained only short syllables (***s***) giving arrhythmic series; the other contained an internal and a final long syllable (***S***) giving the following patterns (where ***S*** is 1.7 times longer than ***s*,** as in French speech).

s S s s s s S
s s S s s s S
s s s S s s S
s s s s S s S
s s s s s S S

The final long ***S*** in these series gave the effect of "group-final" rhythms and all series were adjusted in terms of *P-center* ratios creating natural-sounding rhythms. The stimuli were delivered via a loudspeaker (71 dBA at the Ss' ears). Arrhythmic series were presented first, followed by rhythmic series in random order. The scoring focused on the recall of list-initial syllables with a group mark as opposed to the recall of syllables in the same position but with no rhythm (baseline).

Two results are of interest. (1) Serial recall was significantly superior to baseline for rhythm groups of up to four syllables, but for longer groups, recall was significantly *below* baseline. This suggests that rhythms exceeding four syllables impose a greater load on memory. (2) More importantly, Fig. 3 shows that the decreasing benefit of long groupings on memory appears to correlate with the occurrence of long rhythm groups observed in sentence contexts. Such correspondence suggests a common size principle operating on STM and prosodic grouping in speech.

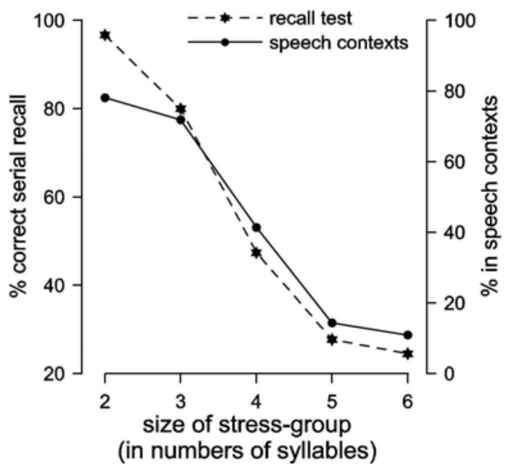

Figure 3: Results of recall and production tests [8].

Several lines of observation support this interpretation. (1) Though recalling grouped series and uttering groups in speech seem unrelated, it is commonly observed that, when recalling lists of syllables or digits like phone numbers, people produce prosodic groups similar to those of speech [8]. Note that such lists do not involve syntax, and yet speakers produce prosody in learning them. This role of prosody may also apply to language acquisition: the fact that children can perceive and produce prosodic patterns before they utter recognizable words (see ref. in [7]) points to the antecedent role of prosody in learning morpho-syntactic forms. (2) As to the generality of size effects, the above results bear on a language where grouping marks are not lexicalized, and it is unclear how this extends to languages with lexical marks like “word” stress. However, inasmuch as a rhythm group contains one stress mark, size effects may be observed. For instance, Dauer [10] found that inter-stress intervals in six languages do not exceed 4.2 syllables on average. But there are also fundamental reasons for assuming that size effects generally apply: all speakers who acquire language face the task of learning rhythmic series of sounds constituting new expressions and, as the above results show, rhythms extending beyond four syllables can have a significant negative impact on serial memory. (3) On the origin of this four-syllable limit, there are important parallels to be drawn with a principle of attention in STM proposed by Cowan [11]. Briefly, Cowan has submitted that people cannot process multiple stimulus channels so there is a *focus of attention*, which can hold up to four “items” or “chunks” at once. Notions of items and chunks are quite controversial, but for verbal series, grouping can be an antecedent condition of chunking, which relates to the formation of units in long-term memory (LTM). For example, Cowan mentions that the way telephone numbers are presented (in groups not exceeding four digits) provides an indication of how many items can be held in the focus of attention at one time to allow the formation of chunks in LTM. It is also suggested that the limit of 7±2 items can reflect that two groups may be held active in STM. Note that “items” here do not reflect words. Well known “word length effects” [12] show that STM capacity operates in terms of numbers of syllables and can be influenced by syllable rate and complexity. At present, our observations show that rhythm and intonation groups are limited to about four and eight syllables respectively [13]. This suggests an intriguing link between prosodic structures and the 7±2 limit on recall of verbal series with possible implications on morpho-syntactic processes.

4. PROSPECTIVE CONCLUSION

Models guided by linguistic concepts often assume segments and words as given, even though heard utterances offer unreliable support for such units. There are concerns that this paradox may rest with linguistic concepts devised by reference to writing. If this is the case, some refocusing of theory will be essential. One advantage of F/C theory is that it proposes working hypotheses of the rise of serial-order and forms where there is no need to assume *a priori* units. However our observations suggest that processes underlying structural aspects of speech do not bear solely on mechanisms of serial-order and motor-sensory coupling, but can also entail constraints on memory processes.

5. REFERENCES

[1] MacNeilage, P.F. 1998. The frame/content theory of evolution of speech production. *Beh. Brain Sc, 21,* 499-546.

[2] Coulmas, F. 1989. *The writing systems of the world* Oxford: Basil Blackwell.

[3] Boucher, V.J. 1994. Alphabet-related biases in psycholinguistic enquiries: Considerations for direct theories of speech production and perception. *J Phonetics, 22,* 1-18

[4] Boucher, V.J. To appear. Intrinsic factors of cyclical motion in speech articulators: Reappraising serial-ordering in motor-control theories. *J Phonetics.*

[5] Ostry, D., Munhall, K. 1985. Control of rate and duration of speech movements. *J. Acoust. Soc. Am,* 77, 640-648.

[6] Laitman, J. T. 1983. The evolution of the hominid upper respiratory system and implications for the origins of speech. In E. De Grolier (ed.), *Glossogenetics: The origin and evolution of language* New York: Harwood, 63-90.

[7] Weissenborn, J., Höle, B. (eds) 2001. *Approaches to bootstrapping. Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition.* Philadelphia: John Benjamin.

[8] 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.

[9] Martin, P. 1999. Intonation of spontaneous speech in French. In *Proc. 15th ICPhS* Barcelona, 17-20.

[10] Dauer, R. M. 1983. Stress-timing and syllable timing reanalyzed. *J Phonetics*, 11, 51-62.

[11] Cowan, N. 2005. *Working memory capacity*. New York: Psychology Press.

[12] Baddeley, A. D., Thomson, N., Buchanan, M. 1975. Word length and the structure of short-term memory. *J Verbal Learn Verbal Beh,* 14, 575-588.

[13] Gilbert, A. C., Boucher, V.J. 2006. Syntax and syllable count as predictors of French tonal groups: Drawing links to memory for prosody. In *Proc 3rd International Conf Speech Prosody*. Dresden, 289-292.