

Phonology and Phonetics

Laboratory Phonology 7

4-1

Editor

Aditi Lahiri

edited by

Carlos Gussenhoven

Natasha Warner

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter
Berlin · New York 2002

Gestural overlap and recoverability: Articulatory evidence from Georgian

Ioana Chitoran, Louis Goldstein and Dani Byrd

Abstract

According to previous investigations of gestural patterning, consonant gestures exhibit less temporal overlap in a syllable/word onset than in a coda or across syllables. Additionally, front-to-back order of place of articulation in stop-stop sequences (labial-coronal, coronal-dorsal, labial-dorsal) exhibits more overlap than the opposite order. One possible account for these differences is that substantial overlap of obstruent gestures may threaten their perceptual recoverability, particularly word/utterance-initially and in a back-to-front sequence. We report here on a magnetometer study of gestural overlap, investigating the role of perceptual recoverability. We focus on Georgian, which allows stop sequences in different positions in the word. C1C2 sequences were examined as a function of position in the word, and the order of place of articulation of C1 and C2. The predictions were borne out: more overlap was allowed in positions where recoverability of C1 is less easily compromised (word-internally and in front-to-back sequences). Similar recoverability requirements are proposed to account for consonant sequencing phenomena violating sonority. Georgian syllable onsets violate sonority, but are apparently sensitive to gestural recoverability requirements as reflected in overlap patterns. We propose that sonority sequencing allows gestures to overlap while still allowing recoverability, but this function can apparently be filled in other ways.

1. Introduction

Linguistic phonetics has long been motivated, or perhaps plagued, by the search for articulatory and acoustic invariance. However, the discovery of invariant aspects of speech has proven elusive. In fact, one of the aspects of speech which has been long recognized as highly variable is its temporal patterning (Gaitenby, 1965; Klatt, 1976). In response to this fact, research has been developed along two directions: (i) pursuing more abstract methods of expressing

timing relations in terms of dynamics and phasing, as part of a general theoretical framework that characterizes the systematic articulatory patterns occurring in speech (Kelso, Saltzman & Tuller, 1986; Browman & Goldstein, 1990; Saltzman & Byrd, 2000); (ii) attempting to identify different linguistic factors which systematically determine surface variation, and to account for the variation in terms of the interaction of multiple constraints, possibly reflecting different levels of phonological structure (Byrd, 1996b; Byrd & Saltzman, 1998). For example, parameter values such as constriction location and degree (traditional, place and manner) of gestures and their position in the larger structures of the syllable, word and phrase have been found to exert systematic influences on gestural timing – both within and between gestural units.

It has also become clear that many of these linguistic effects on the speaker's coordination of gestures are likely to be motivated either diachronically or synchronically (or both) by the need to ensure the best chance at successfully completing the communicative act. That is to say, that the temporal variability shown in speakers' coordination of articulatory movements appears to be sensitive to the necessity of recovering the intended linguistic units from the acoustic signal.

In the present study, we investigate two phonological properties which have been suggested to systematically influence multi-gesture coordination. These are place of articulation and position in the word. These parameters are of particular interest, as the influence they have been hypothesized to exert on intergestural timing seems to reflect the needs of the listener in terms of perceptual recoverability of the coordinated gestures. Specifically, our study focuses on the articulatory patterning of consonant sequences in Georgian, a South Caucasian, Kartvelian language.

Consonants, and particularly the consonants of Georgian, are of special interest for such work. Landmarks in consonant (constriction) articulatory trajectories are reasonably well understood and amenable to analysis. Georgian provides an excellent test bed for questions of intergestural timing as it allows complex sequences of adjacent stop consonants in both word initial and word internal positions – a combination which is rare in the world's languages.

Stop sequences are an important multi-gesture complex to study, as recoverability of these gestures is at particular risk when there is a large amount of gestural overlap. Finally, no articulatory data has been available on Georgian, that we are aware of, apart from X-ray data discussed in Zhgenti (1956).

In this study we use movement tracking data to ascertain the degree of articulatory coproduction (or overlap) in stop-stop sequences as a function of two linguistic factors:

- order of place of articulation (front preceding back versus back preceding front) in the consonant sequences
- position of the consonant sequence in the word (word-initial versus word-internal)

We propose that perceptual recoverability requirements account for consonant sequencing phenomena in Georgian, which violate the sonority sequencing generalization.

We begin in Section 2 with an overview of previous work on the effect of place of articulation and word position on the timing, or gestural overlap, of consonant sequences. The method for the articulatory experiment on Georgian is presented in section 3. Section 4 contains the results, and a discussion is presented in section 5.

2. Background

2.1 Effects of word position

Recent investigations of gestural patterning (e.g., Hardcastle, 1985; Byrd, 1996a) have found that sequences of consonant gestures exhibit less temporal overlap in a word onset than when they occur elsewhere. One possible account for this difference is that substantial overlap of obstruent gestures may threaten their perceptual recoverability, and this may be particularly problematic in utterance-initial position. The proposal that recoverability-related issues guide the coordination of gestures has previously been formulated by Byrd (1994, 1996a, b), Silverman & Jun (1994), Silverman (1995), and Wright (1996). We argue here for a similar approach.

There are two reasons why word onset position might be well protected against a high degree of gestural overlap: both relate to issues of perceptual recoverability. Word onsets are potential utterance onsets and, as such, sequences of stop consonant gestures in this position provide the listener with no acoustic information during their formation. That is, no formant transitions from a preceding vowel into either C1 or C2 are available. Transitions are present only during the release of C2 into a following vowel, but not during the release of C1, since there is no vowel following (for discussion, see Redford & Diehl, 1999). Because the acoustic information for C1 is limited in this way, the degree to which the two consonants may overlap each other might be consequently restricted so as to preserve as much acoustic information as possible about each of the consonants. In this case it becomes crucial for the first consonant in the sequence to be acoustically released if it is a stop, since in word-/utterance-initial position the acoustic release is the only available information as to the presence and nature of that consonant. Furthermore, it is well known that word onsets are important in lexical access (Marsten-Wilson, 1987). This factor might further encourage limits on articulatory overlap in this position to ensure recoverability of the important initial segments.

Several previous studies have contributed to our understanding of the effect of word position on consonant sequence timing. An electropalatographic study of stop-stop and s-stop clusters by Byrd (1996a) indicates that consonant gestures exhibit less temporal overlap in a syllable/word onset than in a coda or across syllables/words. Similarly, for stop-liquid clusters, Hardcastle (1985) found less overlap in onset #k1 than in k#1. Byrd (1996a) also demonstrated that an s-stop sequence occurring as an onset cluster is not only less overlapped, but also less variable in its timing than the same sequence as a coda cluster or a heterosyllabic sequence.

Acoustic data on consonant overlap is available from two languages, Tsou (spoken in Taiwan) and Georgian. Wright (1996) shows acoustic evidence from Tsou stop-stop sequences, suggesting that the timing between articulations is governed by recoverability requirements. In Tsou word-initial stop-stop sequences, a

smaller degree of overlap is allowed than in word-internal sequences. The same is found to be true in an acoustic study of Georgian (Chitoran, 1999). The acoustic signal, however, is not directly informative as to the amount of overlap, nor as to whether the absence of an acoustic release is due to the fact that the respective stop is not released articulatorily, or to the fact that its release is hidden, overlapped by the following stop. It generally can only tell us whether sufficient overlap exists to obscure an acoustic release burst of C1. In both Tsou and Georgian, C1 is always released in word-initial position, and is less systematically released when the stop-stop sequence occurs word-internally. This suggests that overlap is more constrained word-internally.

In summary, the articulatory study we present below will test the hypothesis:

H1: Word-initial stop-stop sequences will be less overlapped than like word-internal sequences.

2.2 Effects of order of place of articulation

Just as perceptual recoverability is argued to be a factor in the word position effect on timing in stop-stop sequences, recoverability considerations may also constrain the patterns of articulatory overlap that occur as a function of the places of articulation of the consonants. In particular, a front-preceding-back order of place of articulation in stop-stop sequences (such as labial-coronal, coronal-dorsal, or labial-dorsal) is expected to allow more overlap than the back-preceding-front order. (In the following we will use the phrase "back-to-front" to refer to a sequence where the more posterior constriction is that of C1 and the more anterior constriction is that of C2; and the reverse for "front-to-back" sequences.) Back-to-front sequences are expected to allow less overlap because, just in the case when the second stop constriction is more anterior than the first, the release of the constriction for the first stop will produce no acoustic manifestation if the constriction for the second consonant is already in place. At a high degree of overlap, the

second constriction lies ahead of the first constriction, which is yet to be released. If, however, the second consonant has a place of articulation more posterior than that of the first, then at least some acoustic information will be generated on release of C1 (even if it does not generate the substantial release burst associated with venting a high-pressure chamber to the atmosphere). Even at a high degree of overlap, because the second constriction lies behind the first constriction, it is less likely to obscure the first consonant's release. The loss of the release information useful in recovering the first consonant in the back-to-front order would be a detriment to perceptual recoverability. Consequently, a more limited degree of overlap would be predicted for such a back-to-front sequence, where C1 gestures are more easily hidden by C2 gestures.

These recoverability considerations may account for several previous experimental results that show that a front-to-back order of place of articulation (labial-coronal, coronal-dorsal, labial-dorsal) allows more overlap than the opposite order, and is more effective in the recoverability of C1 gesture. C1 is more systematically correctly perceived, for example, in labial-coronal than in coronal-labial sequences, even at a higher degree of overlap. Byrd (1992) finds that in the speech stimuli [b#d] and [d#b] synthesized with an articulatory synthesizer, as the amount of overlap is increased, identification of C1 is significantly reduced in [d#b], more so than in [b#d]. This suggests an effect of ordering of the two gestures, and Byrd proposes that a tongue tip gesture is more easily hidden by a following labial gesture than vice-versa. Surprenant & Goldstein (1998) obtained similar results with natural speech [p#t] and [t#p] in English. The tokens used in the perception experiment exhibited the same considerable amount of overlap. C1 in [p#t] was correctly identified significantly more often than C1 in [t#p].

As for other places of articulation, articulatory data from English in Hardcastle & Roach (1979), Zsiga (1994), Byrd (1996a) show that coronal-dorsal sequences ([t#k], [d#g]) allow more overlap than the opposite order ([k#t], [g#d]). Peng (1996) presents results from a perceptual study of place coarticulation in Taiwanese which suggest that a similar overlap pattern is present in this

language (a coronal-dorsal sequence is more overlapped than a dorsal-coronal one). This suggests that a tongue tip gesture is more easily overlapped by a following tongue body gesture than vice versa. However, these studies do not provide articulatory data that include the labial stops.

The effect of order of place is observed in acoustic studies by Wright (1996) for Tsou, and by Chitoran (1999) for Georgian. The acoustic parameter measured in both studies is the inter-burst interval between C1 and C2. The interval was found to be significantly shorter in front-to-back than in back-to-front sequences, suggesting a higher degree of overlap in the former. However, these acoustic findings are again limited in interpretability given that gestural overlap cannot be directly inferred from acoustics alone.

The articulatory movement tracking study reported below evaluates the following hypothesis:

H2: Stop-stop sequences with a back-to-front order of constriction location (coronal-labial, dorsal-labial, dorsal-coronal) will evidence less gestural overlap than stop-stop sequences with a front-to-back order.

We turn in the next section to the description of the experiment, and some information on the phonology of Georgian.

3. Method

Two native speakers of Georgian served as subjects.

3.1 Data

We begin by presenting the stop inventory of Georgian:

(1)	b	d	dz	dʒ	g
	p ^h	t ^h	ts ^h	tʃ ^h	k ^h
	p'	t'	ts'	tʃ'	k'
					q

The stimulus sentences each have a target word containing a stop sequence. The target words were embedded in the frame sentence: *Sit'q'ya — gamoi^hrhinis ordzer*. '*word* — is pronounced twice.' Although more data were collected and analyzed, we report here the results for a subset of the forms, selected for a balanced factorial design. These stimuli are listed in Table 1, with their glosses.

Table 1. Stimuli and glosses in IPA transcription (- indicates a morpheme boundary).

Consonants		Word-initial sequences	Word-internal sequences
C1	C2		
front-to-back			
b	g	bgera	'sound'
ph	t ^h	p ^h t ^h la	'hair lock'
d	g	dg-eb-a	's/he stands up'
back-to-front			
g	b	g-ber-av-s	's/he is inflating you'
t ^h	b	t ^h b-eb-a	'it is warming up'
g	d	gd-eb-a	'to be thrown'

All of the stop-stop sequences are tautomorphic, with the exception of *g-ber-av-s*, a verb form where *g-* is a person marker. The vowels preceding and following the stop sequences are, as often as possible, the low central [a]. This vowel is preferred because it minimally interferes with the trajectory of the consonantal gestures evaluated here. However, not all combinations of consonants occur in lexical items where they are flanked by low central vowels; therefore in some of the stimuli the consonant sequences are followed by the vowels [e], [o], or [i]. The tokens were randomized in stimulus blocks that include other stimuli not analyzed here, and seven repetitions of each block were recorded.

The syllabification of the word-internal stop-stop sequences is not clear, therefore we cannot tell for sure that these sequences span a syllable boundary. Reports from five native speakers (Chitoran, 2000) show that their intuitions on syllabification are very clear only concerning word-initial clusters and single intervocalic consonants. Word-initial clusters are systematically reported to be tautosyllabic, and single intervocalic consonants are syllabified as onsets (e.g. *k'a.lami* 'pen'). Intuitions regarding word-internal clusters are, on the contrary, very mixed. The location of the syllable boundary could not be consistently marked. The only pattern unacceptable to the speakers is *VCC.V, where both consonants are syllabified in the coda of the first syllable, leaving the second syllable onsetless. Morphological boundaries also do not seem to play a role in syllabification, with the exception of compounds, which are not included here.

We should also point out that some of the stop-stop sequences in our list are those referred to in traditional grammars as "harmonic clusters". They are characterized by three properties: (i) the two members of these clusters always share the same laryngeal specifications (voiced, aspirated, or ejective, the three-way laryngeal distinction in the Georgian consonant system); (ii) the order of the place of articulation is always labial-dorsal or coronal-dorsal (e.g. *bg*, *dg*, *t^hk^h*, *ts^hk^h*); (iii) harmonic clusters have been impressionistically described as being single segments, with only one closure and one release. However, acoustic evidence (Chitoran, 1998; McCoy, 1999) indicates that they are sequences of two stops, each with its own closure and release. There is therefore no structural difference between them and the other stop sequences investigated here.

3.2 Data collection

Each frame sentence containing the stimuli was typed in the Georgian alphabet, one sentence per page. The speaker was instructed to read each sentence aloud, at a normal, comfortable pace. The experimenter cued the speaker for each sentence by the word "Go." If the speaker paused or had a false start, he was asked to

re-read the sentence. The 12 stimuli were read 7 times, therefore a total of 84 stimuli were recorded. In spite of the careful data collection, a few stimuli were lost due to technical problems (e.g. transducers came loose). Thus, three stimuli were lost from the first speaker (one of each: *p^hi^hila*, *ap^hi^hari*, *abga*), and four from the second speaker (one of each: *dgeba*, *galeba*, *p^hi^hila*, *adgēba*).

Data were collected using the EMMA (Electromagnetic Midsagittal Articulator) magnetometer system. The technical specifications of the EMMA magnetometer system are outlined in Per-kell, Cohen, Svirsky, Garabietta & Jackson (1992) (see also Gracco & Nye, 1993; Löfqvist, 1993). Receivers were attached to three midsagittal points on the subject's tongue. One, (TT) was positioned as posterior as possible, another (TU) was attached approximately 1 cm from the tongue tip, and a third was positioned at an intermediate location. In addition, receivers were placed on the upper and lower lip, the lower and upper teeth (maxilla & jaw, respectively) and the nose bridge, the latter two for correction of head movement. Both acoustic and movement data were obtained. The movement data were sampled at 500Hz (after low-pass filtering before voltage-to-distance conversion) and the acoustic data were sampled at 20kHz. The data were corrected for head movement and rotated to the occlusal plane of the subject such that the x-axis is parallel to the occlusal plane and the y-axis lies perpendicular to it. Voltages were low-pass filtered at 15 Hz, using a 9th-order Butterworth filter. After voltage to distance conversion, correction for head movement (using the nose and maxillary reference transducers), and rotation to the occlusal plane, the position signals were also low-pass filtered at 15 Hz.

Movement trajectories of the receivers attached to the tongue tip (TT), tongue dorsum (TD), upper lip (UL) and lower lip (LL) were evaluated. For coronal stops ([t^h], [d]) the tangential velocity (xy) minima of the tongue tip receiver were used to delimit the gestures' temporal location. This was calculated as follows:

$$tvel = \sqrt{(V_x)^2 + (V_y)^2}$$

where: *tvel* = tangential velocity of the tongue tip transducer

V_x = velocity in the x-coordinate of the tongue tip transducer

V_y = velocity in the y-coordinate of the tongue tip transducer

For velar stops ([g]) the velocity zero-crossings of the vertical (y) movement of the tongue dorsum receiver was employed. For labial stops ([b], [p^h]) both the upper lip and lower lip receiver vertical (y) trajectories were evaluated. Both speakers showed considerable displacement of the upper lip during the closure phase of bilabial stops, and of the lower lip during the release phase. We therefore employed the movement of the two lips separately, rather than as a single variable (lip aperture), representing the vertical distance between the two lips. Upper lip velocity zero-crossings were used to identify the onset and achievement of the labial constriction, and lower lip movement was used to identify the constriction release.

3.3 Analysis

The data were analyzed using MATLAB to algorithmically identify important landmarks in the movement trajectories. For each gesture the following three points were identified and labeled: movement onset (labeled *On*), target achievement (point at which constriction is achieved, labeled *Off*), and target release (point at which constriction is released, labeled *On*). Onsets of motion were defined algorithmically as the points in time at which the velocity exceeded some specified threshold above zero velocity. Offsets were defined as the points where velocity fell below that same threshold. Thresholds were set as a percentage of the effective maximum speed that each receiver dimension exhibited over all utterances. The effective maximum speed was calculated by finding the maximum speed (absolute value of velocity) observed in the middle 1/3 of each utterance, and then averaging across all utterances. Percentages were as follows: for TT, 15% of the effective maximum tangential velocity (xy); for TD, 15% of the effective maximum vertical (y) speed; for UL, 20% of the effective maximum vertical (y) speed to identify the onset of movement and the constriction

achievement; for LL, 15% of the effective maximum vertical (y) speed to determine the constriction release.

As an index of the temporal overlap between the two sequential stop gestures, the following measure was evaluated in the quantitative analyses: the percentage of the interval between target achievement and release for the first stop at which movement onset for the second stop is initiated. That is, how early does C2 movement onset occur within the constriction 'plateau' interval of C1? This measure will be referred to as overlap. A small value indicates a large degree of overlap, that is, C2 movement starts quite early in C1's constriction, and a large value indicates little or no (if > 100%) overlap, that is, C2 movement starts quite late in C1's constriction, or after it. A sample data panel is shown in Figure 1:

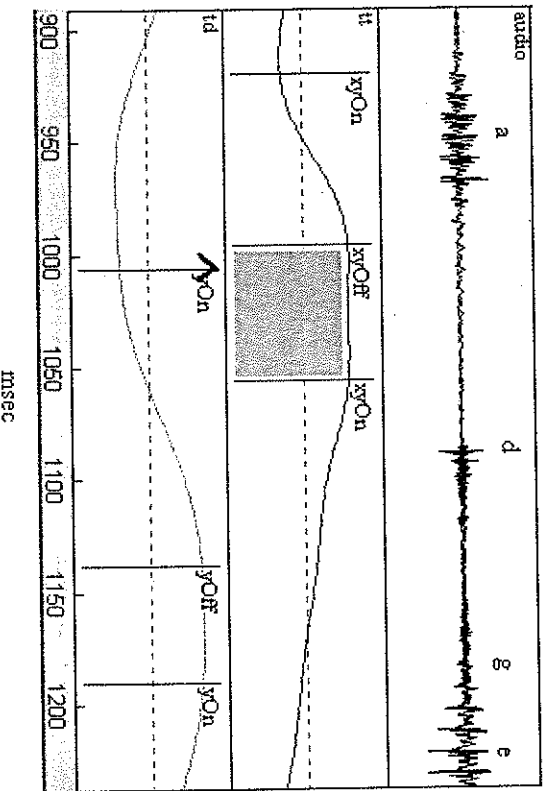


Figure 1: The sequence [a#d#e] in [sɪt'q'vɑ#dʒeɪl]. Top panel: audio; Middle panel: tongue tip y with events labeled; Bottom panel: tongue dorsum y with events labeled. Labels: On indicates onset of movement, toward and away from target. Off indicates offset of movement, target achievement. Note that in this token C2 initiates shortly after C1 constriction is achieved.

4. Results

The overlap measure is evaluated as the dependent variable in a 3-factor full-interaction ANOVA model for each speaker separately. The three independent variables for the non-repeated measures ANOVA are:

- (i) POSITION (2 levels: word-initial, word-internal)
- (ii) ORDER (2 levels: front-to-back, back-to-front)
- (iii) PLACES (3 levels: labial & coronal, coronal & dorsal, labial & dorsal; irrespective of order of occurrence)

Table 2: Summary of results (significant effects in bold).

Effect	Speaker 1	Speaker 2
POSITION	F(1,69)=51.48, p<.001	F(1,68)=.06, p>.05
ORDER	F(1,69)=52.91, p<.001	F(2,68)=8.37, p<.01
PLACES	F(2,69)=17.03, p<.001	F(2,68)=8.63, p<.001
POSITION x ORDER	F(1,69)=.23, p>.05	F(1,68)=8.07, p<.01
POSITION x PLACES	F(2,69)=.81, p>.05	F(2,68)=3.41, p>.01
ORDER x PLACES	F(2,69)=.16, p>.05	F(2,68)=7.95, p<.01
POSITION x ORDER x PLACES	F(2,69)=1.5, p>.05	F(2,68)=3.56, p>.01

Table 2 shows that all main effects are significant, with the exception of the POSITION effect for the second speaker. No two-way or three-way interactions are significant for the first speaker. For the second speaker, two two-way interactions are significant: POSITION x ORDER and ORDER x PLACES. While the first interaction, to be discussed below, is relevant to our hypotheses, the second one is not. We did not predict an effect of PLACES, and we do not have at present an explanation for this significant interaction in the case of only one speaker.

4.1 Effects of POSITION

An effect of POSITION was found for both speakers, with some individual differences. The data of the first speaker show a significant effect of POSITION, and no significant interactions. The second speaker shows no main effect of POSITION, but the interaction between POSITION and ORDER is significant.

The main effect of POSITION obtained for the first speaker indicates that stop-stop sequences in word-internal position have a significantly greater amount of overlap than like sequences in word-initial position. This finding confirms hypothesis H1 and is consistent with similar findings for English outlined in section 2.1. Furthermore, the results also show that in word-internal sequences, C2 onset occurs on average soon after the achievement of C1 target, after only 5% of the C1 constriction interval, whereas in word-initial sequences C2 onset occurs much later (after an average of 82% of the interval). The results for the first speaker are summarized in Table 3, and illustrated in Figure 2, further below.

While the lack of a main effect of position for the second speaker fails to support H1 as it is stated, the pattern of this speaker's results is consistent with the reasoning behind H1. The results show that in word-initial position, where release is hypothesized to be critical to recoverability, front-to-back sequences are more overlapped than back-to-front ones (as predicted by H2). In the word-initial front-to-back stop sequences, C2 onset occurs on average even before the constriction for C1 is achieved (-42% of the C1 constriction interval). In back-to-front sequences, on the other hand, there is a long delay for the C2 onset, which occurs on average after 62% of the C1 constriction interval. Word-medially, however, where more information is available to support recoverability, the amount of overlap is comparable, as discussed further in section 4.2. The results for the second speaker are summarized in Table 4, and illustrated graphically in Figure 3, further below.

4.2 Effects of ORDER

The data of the first speaker show a significant main effect of ORDER, and no significant interactions. In the front-to-back stop sequences, C2 onset occurs on average after 3% of the C1 constriction interval, as opposed to back-to-front sequences, where C2 onset occurs after 82% of the interval. The second speaker also shows a significant main effect of ORDER, but, as discussed above, a significant interaction is also found between POSITION and OR-

DER. The simple main effect of ORDER is significant in word-initial position, but not word-medially. For both orders of place, C2 onset occurs on average after 17% of the C1 constriction interval.

Overall, these results confirm our hypothesis H2: stop-stop sequences with a front-to-back ordering of place of articulation show significantly more overlap than sequences with the reversed order of place of articulation. This result is also consistent with previous findings summarized in section 2.2.

4.3 Effects of PLACES

Finally, a significant main effect of PLACES (the types of gestures involved - labial, coronal or dorsal) was found for both speakers. This effect was not one that we had predicted. For the first speaker the highest degree of overlap was found for dorsal-coronal/coronal-dorsal sequences, where C2 onset occurs on average after only 4% of the C1 constriction interval. The next most overlapped sequences are labial-dorsal/dorsal-labial. C2 onset occurs on average after 46% of the C1 constriction interval. The least overlapped sequences are labial-coronal/coronal-labial. C2 onset occurs on average after 85% of the C1 target-release interval.

For the second speaker, as for the first, the least overlapped sequences are those combining labials and coronals. C2 onset occurs on average after 60% of the C1 constriction interval. The other two place combinations have higher degrees of overlap: 24% for labial-dorsal/dorsal-labial, and 7% for coronal-dorsal/dorsal-coronal.

4.4 Summary of results by speaker

The results for each speaker are summarized in Tables 3 and 4, respectively. Those specific to Hypotheses 1 and 2 are shown graphically in Figures 2 and 3, for each speaker. The results will be further discussed in section 5.

Table 3: Summary of overlap means for Speaker 1 (C2 onset relative to C1 constriction interval). Lower numbers indicate greater overlap.

	n	Mean (%)	S.D. (%)
POSITION			
word-initial	41	82%	75
word-internal	40	5%	63
ORDER			
front-to-back	39	3%	54
back-to-front	42	82%	80
PLACES			
labial & coronal	26	85%	84
labial & dorsal	27	46%	75
coronal & dorsal	28	4%	59
POSITION x ORDER (ns)			
word-initial			
front-to-back	20	39%	31
back-to-front	21	124%	82
word-internal			
front-to-back	19	-34%	47
back-to-front	21	41%	54

The interaction between POSITION and ORDER is plotted in the graph in Figure 2, below.

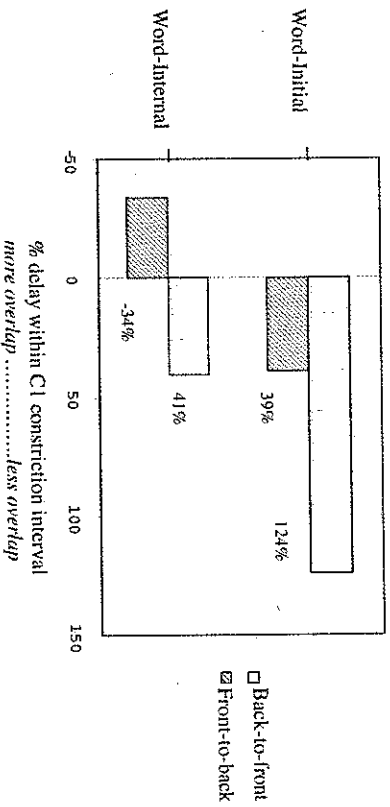


Figure 2: Amount of overlap in C1 constriction interval (speaker 1)

The graph indicates very little overlap in the back-to-front word-initial sequences, as expected from the main effects. A mean overlap measure greater than 100% indicates C2 onset occurs on average after the release of the C1 constriction. The most overlapped sequences are the front-to-back word-internal sequences. A negative mean overlap measure indicates C2 onset occurs before the target achievement for the C1 constriction.

Table 4: Summary of overlap means for Speaker 2 (C2 onset relative to C1 constriction interval)

	n	Mean (%)	S.D. (%)
POSITION (ns)			
word-initial	39	10%	130
word-internal	41	17%	52
ORDER			
front-to-back	39	-12%	127
back-to-front	41	39%	47
PLACES			
labial & coronal	27	60%	37
coronal & dorsal	25	7%	48
labial & dorsal	28	-24%	145
POSITION x ORDER			
word-initial			
front-to-back	19	-42%	172
back-to-front	20	62%	21
word-internal (ns)			
front-to-back	20	17%	52
back-to-front	21	17%	54

The interaction between POSITION and ORDER is plotted in the graph in Figure 3. The graph indicates that a much greater amount of overlap is allowed in word-initial front-to-back sequences than in back-to-front ones, in keeping with Hypothesis 2. In word-internal position, however, the same amount of overlap is found, regardless of order of place.

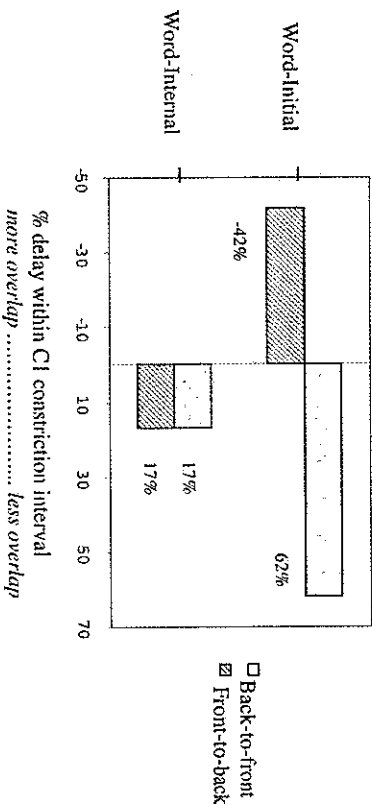


Figure 3: Amount of overlap in C1 constriction interval (speaker 2)

5. Discussion

5.1 Patterns of gestural overlap

The results of this movement tracking study of the articulation of Georgian consonant sequences confirm both of the original hypotheses regarding degree of overlap as a function of position in the word and of order of place of articulation. Specifically:

- More gestural overlap is found word-internally than word-initially (H1);
- More gestural overlap is found in sequences with a front-to-back order of place of articulation than in back-to-front sequences (H2).

The data of the first speaker supports both hypotheses: word position and constriction order both affect the relative timing among gestures. The data of the second speaker confirm the order hypothesis, but only in word-initial position. For both speakers, an additional, unpredicted effect of place combination was also found, such that combinations of labial and coronal stops are the least overlapped. The PLACES effect can be readily interpreted, however, since labial and coronal sequences are in fact the ones which are not usually described as “harmonic” in Georgian, and not claimed

to be single segments. It is possible that their relatively reduced degree of overlap may in fact be responsible for their perception as a clear sequence of consonants. It is also significant to note that cross-linguistically, double articulations of labials and coronals are not attested. Closure gestures can be nearly synchronous in combinations of labials and velars, and in combinations of coronals and velars in clicks. Labials and coronals, though, do not seem to display a high degree of overlap. The implications of degree of overlap for the status of “harmonic” clusters in Georgian will be further discussed in section 5.3.

Our original hypotheses concerning word position and order of place of articulation were both motivated by the view that listeners’ needs for perceptual recoverability play a role in determining the spatiotemporal patterning of gestures produced by speakers (Matingly, 1981; Silverman, 1995). Less gestural overlap between consonants in sequences helps preserve the information that serves to specify the identity of the first consonant, in particular the release burst of C1 and the specificity of the VC and CV formant transitions for each of the consonants. With regard to the gestural overlap (or phase relations) among gestures, the experiment results provide preliminary evidence that timing patterns reflect constraints of perceptual recoverability in at least two different respects. First, less overlap (or better preservation of information on consonant identity) is found in a prosodic position which is critical for lexical access – the word onset. Second, less overlap is observed when C1 is especially vulnerable to being obscured by C2.

Of course, recoverability, while a primary concern for both speaker and listener, is not the only influence on spatio-temporal variability in speech. Efficiency is a parallel concern. Gestural overlap or coproduction presents an advantage from the point of view of transmitting information simultaneously about several linguistic units – what has been called *parallel transmission* (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). Thus, it seems that there are competing influences on intergestural timing: the first is the need to ensure recoverability of linguistic units from the signal, and the second is the need to encode and transmit information at a high rate.

Theories of speech production must have some way of incorporating these competing influences on speech timing. One approach is outlined in Byrd (1996b) in which she views influences on phase relations as interacting probabilistically. Another approach has been offered by Browman & Goldstein (2000). They note that gestural parameters and phase relations may be quantitatively scaled as a function of speaking conditions. One step toward accomplishing this is to allow phase relations between gestures to have different degrees of cohesion or bonding strength (Browman & Goldstein, 2000). Sources of variation in gestural overlap influence a given pair of gestures in inverse proportion to their bonding strength. The surfacing temporal pattern is taken to be the result of competing phase relations such that the surface timing maximizes the satisfaction of the competing constraints as weighted by their bonding strength (Browman & Goldstein, 2000). The specific example sketched by Browman & Goldstein in their presentation includes a C-V relation that defines syllable onsets, and a C-C relation between two adjacent consonants. According to the C-V relation, each consonant gesture in a complex onset bears the exact same phase relation to the vowel. If each consonant gesture is coordinated in the same way with respect to the nuclear vowel gesture, the consonants tend to synchronize. However, the second relation (the C-C relation) phases the consonant gestures with respect to each other in a way which allows them to be recoverable. According to this relation, the consonants tend to be sequential. Browman and Goldstein describe these competing constraints as being able to account for the observation that in #CC(C)V sequences the temporal center of the consonant interval ("C-center") maintains a fixed temporal relation to the vowel. In more general terms, they view these constraints as characterizing the desirability of simultaneous parallel transmission (the C-V relation) and of perceptual recoverability (the C-C relation). These two tendencies are both accommodated in speech, and furthermore, can be explicitly incorporated into their model of intergestural timing.

Within this model, the results of our study can be interpreted as follows. For both speakers, the bonding strength of the C-C

relation varies as a function of ORDER and POSITION. For the first speaker they combine linearly, giving the independent contribution of the two factors. The data of the second speaker can be explained with the same kinds of constraints, but simply under the assumption that the word-medial consonants are not tautosyllabic. If this is the case, the C-V constraint does not come into play in this context. The remaining C-C constraint does not compete against anything, and thus differences in weights do not have any effect on the observed timing patterns.

If this analysis is correct, it makes a number of predictions about the C-center relations: both speakers should show a C-center effect in word-initial position (we have vowel duration measurements showing that this is the case for the second speaker), but word-medially only the first speaker should show a C-center effect. The results are in fact consistent with the native speakers' mixed intuitions about syllabification in word-medial clusters.

5.2 Implications for sonority sequencing violations

The two potentially opposing tendencies presented above have implications for understanding consonant sequencing in Georgian and other languages that violate the sonority sequencing principle. The sonority principle states the cross-linguistic tendency for complex onsets to rise in sonority toward the syllable nucleus, and complex codas to fall in sonority away from the syllable nucleus. The sonority scale based on this principle contains vowels at the most sonorous end, and obstruents at the least sonorous end:

Obstruents < Nasals < Liquids < Glides < Vowels

The sonority principle is thus primarily a generalization capturing the observation that certain types of onset consonant sequences are the most common cross-linguistically. No consistent phonetic (acoustic or articulatory) correlate of sonority exists, but most attempts to define sonority phonetically relate it to the notion of increased perceptibility of segments. Some of the definitions that we find most intuitive (Matingly, 1981; Ohala, 1990) refer to per-

ceptibility as the ease with which individual segments or gestures are correctly identified in a sequence.

Matingly's (1981) proposal is based on the notion of parallel transmission as an important organizing principle for speech communication. Phonetic elements, although perceived as ordered, are not produced in strict succession. There is a clear many-to-many relationship between phonological units and acoustic cues to these units. The simultaneous availability of information for multiple segments is claimed to facilitate higher information rates for speech. Sonority is one way of achieving this goal of parallel transmission. He treats the traditional sonority ranking as a ranking of manner classes according to degree of "closeness." The scale is argued to follow an ordering that crucially depends on the degree to which information is available during the release or application of the constriction, and during the constriction itself.

"The general articulatory prerequisite for parallel transmission would appear to be that the constriction for one or more closer articulations must be in the process of being released or applied in the presence of constrictions for one or more less close articulations. In terms of this formulation, the conventional ranking of manner classes... corresponds to a ranking according to the degree to which information can be encoded during the release or application of the constriction, and the inverse of this ordering, to the degree to which information can be encoded during the period of maximal constriction... [T]he articulations of speech must be scheduled so that periods during which constrictions are released in rank order alternate with periods during which constrictions are applied in inverse rank order. This is of course exactly what is accomplished by the syllabic organization of speech." [Matingly, 1981, p. 418]

This view does indeed define a ranking very similar to the sonority scale, if we consider how much and what kind of acoustic information is available from different segments. For stops (both oral and nasal), acoustic information for place is present primarily at the application and release of the constriction, and only manner information is present during the closure interval. In the case of liquids, place information is also available during the constriction itself. The same is true of fricatives, although perhaps less so of non-sibilant fricatives, which have lower energy. For glides and vowels, both place and manner information is present throughout the constriction formation and release process. In a complex onset struc-

tured according to this scale, articulations are released into more open constrictions, thereby allowing a relatively high degree of overlap. Overlap in turn allows more acoustic information to be transmitted for more than one phonetic element at the same time. This provides an elegant foundation for the generalization captured traditionally by the sonority scale and, more generally, for the syllabic organization of speech.

Such an approach to sonority is also consistent with the views held by Ohala & Kawasaki (1984) and Ohala (1990), who propose that the salience of an acoustic signal may be given by maximal modulations in several acoustic parameters varying simultaneously (e.g. amplitude, periodicity, spectral shape, fundamental frequency). Preferred sequences of segments would be characterized by large modulations in acoustic parameters. The more acoustic information is present simultaneously, the more successful the identification of the component segments is by the listener. For these modulations to be available to the listener, however, the sequencing must also follow the patterns of intergestural timing we proposed, which best ensure that the relevant acoustic information will not be obscured.

We would like to suggest that such patterns of gestural coordination that satisfy both parallel transmission of information and recoverability are "attractors" towards which phonological structures evolve (through talker-listener interaction), and that these "attractors" underlie the traditional sonority principle. Stop-liquid sequences, for example, are so common cross-linguistically because they allow substantial overlap while maintaining recoverability. A number of languages evolve uncommon sequences ("sonority plateau" or "reversals"). These less common sequences are not ideal for the transmission of gestural information, unless their gestural timing is tightly controlled. We have seen that Georgian syllable onsets violate the sonority sequencing generalization, but appear to be sensitive to gestural recoverability requirements as reflected in overlap patterns. The articulatory patterning of Georgian onsets can be explained by the same opposing tendencies at work in the more common patterns of sonority sequencing: the tendency for gestures to overlap so as to allow parallel transmission, and the

tendency to limit the amount of overlap in order to ensure perceptual recoverability. Languages which obey sonority sequencing may also demonstrate effects of perceptual recoverability on intergestural timing.

We propose, therefore, that cross-linguistically syllable structure adheres to a particular ordering or temporal patterning that allows for maximum overlap with minimal loss of information. We would expect the more common sequences to exhibit relatively less sensitivity to gestural coordination. For example, a velar-liquid onset cluster should exhibit a comparable place effect to a bilabial-liquid onset cluster. The order effect should be negligible here, because acoustic release of either stop is equally unlikely to be obscured by the following liquid, produced with a more open vocal tract. On the other hand, in sonority plateaus and sonority reversals, such as the sequences found in Georgian, intergestural timing is expected to play a crucial role, and the magnitude of the order of place effect should be greater.

5.3 Implications for the status of "harmonic" clusters

The account presented here can be extended to explain the impressionistic descriptions of "harmonic" clusters in Georgian as being single segments. Recall that "harmonic" clusters are front-to-back C1C2 sequences where C2 is dorsal, and C1 and C2 share the same laryngeal specification (i.e. voiced, voiceless, or ejective). Such sequences are common in Georgian. Of the four logically possible combinations of consonants in terms of order of place and laryngeal homogeneity, three are attested in the language (the third being the "harmonic" type):

- homogeneous, back-to-front: gdeba 'to throw'
- non-homogeneous, back-to-front: q'ba 'jaw'
- homogeneous, front-to-back: dgoma 'standing'
- non-homogeneous, front-to-back: *unattested* (*d^hk', *p^hk', *p'g)

We argue that the large amount of overlap in front-to-back sequences is responsible for the absence of front-to-back sequences that do not agree in laryngeal features.

We hypothesize that in Georgian a consonant cluster licenses at most a single laryngeal gesture: glottal abduction (for aspirated consonants) or a laryngeal raising/closing gesture (for ejectives). Voiced stops are hypothesized to result from no active laryngeal maneuver at all (i.e. the speech-default position of the vocal folds is that of adduction appropriate for vibration). Voiced obstruents in Georgian have very weak voicing, and are not necessarily voiced throughout the closure.

We also hypothesize that in Georgian the laryngeal gesture, if present, is coordinated in such a way that its target (opening or closing) is achieved during the first of the two stops. This means that in the case of the more overlapped front-to-back sequences the time between the laryngeal event (coordinated with the first consonant) and the second consonant will be relatively short. Therefore, the laryngeal gesture will still have its characteristic effect (opening or closure) at the time of the release of the second consonant. This would explain why front-to-back clusters *in particular* always agree in voicing since the laryngeal gesture extends through both of them due to their high degree of overlap.

If any laryngeally heterogeneous sequences ever existed, they would be predicted to undergo voicing assimilation. A proposal by Gankrelidze & Ivanov (1995) supports this prediction from a diachronic perspective. They argue that harmonic clusters in Proto-Kartvelian were complex consonants that later segmented into clusters. (They were probably velarized consonants, which would explain why in the modern languages the second consonant is always dorsal, a velar or uvular stop or fricative.) The authors suggest that non-homogeneous, front-to-back sequences did exist, but have undergone voicing assimilation, merging with "harmonic" clusters.

As for the non-harmonic back-to-front sequences that have less overlap, there will be a greater delay between the laryngeal gesture (coordinated with the first consonant) and the release of the second consonant. This accounts for the mixed voicing found in back-to-front (non-harmonic) sequences. This delay may be sufficient for the larynx to return to its default state, resulting in a weakly voiced or voiceless unaspirated stop. This is in fact what is observed for C2 in such clusters (Chitoran, 1999).

Of course, this account of Georgian harmonic clusters still remains to be tested, ideally with instrumentation that will allow us to evaluate laryngeal behavior and its coordination with supralaryngeal articulation.

6. Conclusions

This study contributes to our understanding of cross-linguistic patterns of gestureal overlap, and of the types of factors influencing them. We tested the effects of two linguistic factors (position in the word and order of place of articulation) on the amount of overlap allowed between consonantal gestures in Georgian stop sequences. Both were found to systematically affect the temporal coordination of linguistic units. We discussed the implications of the results for the theoretical conceptualization of articulatory timing. We proposed to account for the results in terms of weighting of constraints that ensures the perceptual recoverability of gestures, while allowing efficiency through overlap.

We outlined a proposal as to how sonority sequencing may evolve from the same competing constraints, reflecting the desirability of efficient patterns of information transmission to listeners while ensuring that perceptual recoverability is possible for the listener.

Finally, we argued that the substantial amount of overlap found in Georgian front-to-back stop sequences may be responsible for the so-called "harmonic" clusters in Georgian, in which the two members of the clusters always agree in laryngeal specification.

Acknowledgements

The authors gratefully acknowledge the support of NIH grant DC-03172 and NSF grant SBR-951730. We would like to thank the LabPhon7 participants, the reviewer, and Alice Harris for their useful comments.

References

- Browman, C. & Goldstein, L.
1990 Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, (pp. 341–376). Cambridge: Cambridge University Press.
- Browman, C. & Goldstein, L.
in press Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5, 24–34.
- Byrd, D.
1992 Perception of assimilation in consonant clusters: a gestural model. *Phonetica*, 49, 1–24.
- Byrd, D.
1994 *Articulatory Timing in English Consonant Sequences*. PhD dissertation, UCLA, published as UCLA Working Papers in Phonetics 86.
- Byrd, D.
1996a Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209–244.
- Byrd, D.
1996b A phase window framework for articulatory timing. *Phonology*, 13 Vol 2, 139–169.
- Byrd, D. & Saltzman, E.
1998 Intergestural dynamics of multiple phrasal boundaries. *Journal of Phonetics*, 26, 173–199.
- Chitoran, I.
1998 Georgian harmonic clusters: phonetic cues to phonological representation. *Phonology*, 15, 121–141.
- Chitoran, I.
1999 Accounting for sonority violations: the case of Georgian consonant sequencing. *Proceedings of the XIV International Congress of Phonetic Sciences*, San Francisco, 1999, 101–104.
- Chitoran, I.
2000 Some evidence for feature specification constraints on Georgian consonant sequencing. In O. Fujimura, B. Joseph & B. Palek (eds.), *Proceedings of LP '98*, 185–204.
- Gaitanby, J.
1965 The elastic word. *Haskins Laboratories Status Report on Speech Research*, 2.
- Gamkrelidze, T. V. & Ivanov, V.
1995 *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a Proto-Language and a Proto-Culture*. (English version by Johanna Nichols). Berlin, New York: Mouton de Gruyter.

- Gracco, V. L. & Nye, P. W.
1993 Magnetometry in speech articulation research: Some misadventures on the road to enlightenment. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, 31.
- Hardcastle, W. J.
1985 Some phonetic and syntactic constraints on lingual coarticulation during /k/ sequences. *Speech Communication*, 4, 247–263.
- Hardcastle, W. J. & Roach, P.
1979 An instrumental investigation of coarticulation in stop consonant sequences. In H. Hollen & P. Hollen (eds.), *Current Issues in the Phonetic Sciences*, (pp. 531–540). Amsterdam: John Benjamins.
- Kelso, J. A. S., Saltzman, E. & Tuller, B.
1986 The dynamical perspective on speech production: data and theory. *Journal of Phonetics*, 14, 29–59.
- Klatt, D. H.
1976 Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1221.
- Lieberman, A., Cooper, F. S., Shankweiler, D. & Studdert-Kennedy, M.
1967 Perception of the speech code. *Psychological Review*, 74, 431–461.
- Löfgqvist, A.
1993 Electromagnetic transduction techniques in the study of speech motor control. Reports from the Department of Phonetics, University of Umeå, PHONUM, 2, 87–106.
- Marslen-Wilson, W. D.
1987 Functional Parallelism in Spoken Word-Recognition. *Cognition*, 25, 71–102.
- Matingly, I. G.
1981 Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver & J. Anderson (eds.), *The Cognitive Representation of Speech*, (pp. 415–420). North Holland Publishing Company.
- McCoy, P.
1999 Harmony and sonority in Georgian. In *Proceedings of the XIV International Congress of Phonetic Sciences*, San Francisco, 1999, 447–450.
- Ohala, J. J.
1990 Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Chicago Linguistic Society*, 26, Vol 2, *The parassession on the syllable in phonetics and phonology*, 319–338.
- Ohala, J. J. & Kawasaki, H.
1984 Prosodic phonology and phonetics. *Phonology Yearbook*, 1, 113–127.
- Peng, Shu-Hui
1996 *Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi*. PhD dissertation, Ohio State University.
- Petkell, J., Cohen, M., Swirsky, M., Matthies, M., Garabeta, I. & Jackson, M.
1992 Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92, 3078–3096.
- Redford, M. A. & Diehl, R. L.
1999 The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America*, 106, 1555–1565.
- Saltzman, E. & Byrd, D.
2000 Demonstrating effects of parameter dynamics on gestural timing. May 2000 Meeting of the ASA, Atlanta, GA. *Journal of the Acoustical Society of America*, 107, Vol 5, 2, 2904.
- Silverman, D. & Jun, J.
1994 Aerodynamic evidence for articulatory overlap in Korean. *Phonetica*, 51, 210–220.
- Silverman, D.
1995 Phrasing and Recoverability. UCLA Ph.D dissertation, published by Garland Publishing, 1997.
- Surprenant, A. M. & Goldstein, L.
1998 The perception of speech gestures. *Journal of the Acoustical Society of America*, 104, Vol 1, 518–529.
- Wright, R.
1996 Consonant clusters and cue preservation in Tsou. PhD dissertation, UCLA. Dissertations in Linguistics, 20.
- Zhgenti, S.
1956 *Kartuli ensi ponetika (Phonetics of the Georgian language)*. Tbilisi.
- Zsiga, E. C.
1994 Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics*, 22, 121–140.