

## Timing lag matters in native speakers' perception of Georgian stop sequences

**Ioana Chitoran\***, Université Paris Cité, CNRS, Laboratoire de linguistique formelle, France; Institut Universitaire de France, France, [ioana.chitoran@u-paris.fr](mailto:ioana.chitoran@u-paris.fr)

**Harim Kwon**, Seoul National University, South Korea

\*Corresponding author.

---

This study tests the hypothesis that timing lag is part of phonological knowledge, investigating Georgian speakers' perceptual recoverability of C1 consonantal gestures in stop-stop sequences. We hypothesize that the recovery of C1 is facilitated by (i) longer timing lag (i.e., reduced gestural overlap) between C1 and C2, and (ii) the presence of a C1 vocalic release. Two perception experiments (a forced choice identification and a follow-up transcription task) were conducted using stimuli that have been acoustically and articulatorily analyzed. We consider three acoustic parameters (acoustic lag, presence of C1 vocalic releases, vocalic release duration) and three articulatory parameters derived from EMA data (release lag, onset lag, relative overlap). The results show that longer timing lag does facilitate the accurate recovery of C1, but the presence or longer duration of a C1 vocalic release does not. The results support the inclusion of timing lag in the phonological grammar. We discuss why vocalic releases provide less reliable perceptual information that is potentially ambiguous between segmental and prosodic (syllabic) levels of structure.

---



## 1. Introduction

The accurate perception of phonetic information that encodes phonological contrasts can be inhibited or enhanced in certain contexts. Misparsings of the speech signal can result from such misperceptions and may ultimately lead to sound change (Ohala, 1981; Harrington et al., 2008; Harrington et al., 2019; Beddor, 2009; Iskarous & Kavitskaya, 2018, among others). It has also been proposed that knowledge of such perceptual patterns is directly encoded in phonological representations (see Hayes et al., 2004; Steriade, 2008). Testing the status of perceptual knowledge in synchronic phonology requires evidence that particular structures are either preferred or avoided, depending on the recoverability of essential parts of the signal. While several previous studies have examined such preferences in production, comparatively fewer have addressed this question directly from perception: Do the preferred structures truly present an advantage for the recoverability of phonological information? The current study contributes to further understanding of this issue empirically and considers its implications for a model of phonological representations.

One path toward testing the role of perceptual information in phonology is shaped by the concept of perceptual recoverability, as developed in articulatory phonology (Browman & Goldstein, 1992; Goldstein & Fowler, 2003), and interpreted with respect to the relative timing of gestures. Perceptual recoverability is understood to be one of the factors that characterizes intergestural coordination (cf. Mattingly, 1981), in the sense that speakers acquire knowledge of the temporal patterns and their acoustic consequences needed to correctly recover a gesture. Mattingly (1981) has proposed, for example, that syllabic organization may reflect such knowledge, because it allows considerable overlap of adjacent gestures, thus maximizing parallel transmission of information, while still allowing the individual gestures to be correctly recovered. When the degree of overlap exceeds the recoverability limits, a speech gesture may be hidden, as shown in the classic example of the phrase *perfect memory* (Browman & Goldstein, 1990). In a fluent production of this phrase, the final /t/ in *perfect* is articulatorily present, but can be acoustically masked by the lip closure of /m/ in *memory* due to the maximum overlap of the consonantal gestures. This shows that even fully articulated gestures may become perceptually inaccessible, highlighting the importance of timing for perceptual recoverability. In articulatory phonology (henceforth AP), variability in inter-gestural timing follows from the dynamic nature of a gesture. A gesture is a dynamical system, and its dynamic regime is affected by a number of factors (see, for example, Sorensen & Gafos, 2016; Iskarous, 2017; Du & Gafos, 2022). In the specific case of the perceptibility of stop-stop sequences, relevant factors that have been examined include position in the word or utterance, and the order of the gestures' respective constriction locations (see references reviewed in Section 2). Both of these factors can be responsible for cases of maximum overlap where one of the gestures can be obscured, and both have been tested in experimental studies that adopt an AP position.

In our study, we further probe to what extent, and under which conditions, timing affects perceptual recoverability of gestures in consonantal sequences. Specifically, we test a hypothesis on the recoverability of stop-stop sequences: Gestures in a C1C2 stop sequence will be harder to recover if the sequence is produced with increased temporal overlap. C1, the first stop in the sequence, is expected to be the most vulnerable perceptually, if its release is overlapped with a complete oral closure. Thus, when a stop-stop sequence is produced with short lag (i.e., increased temporal overlap), the C2 closure can mask the acoustic release of C1. Cross-linguistic evidence in production (reviewed in Section 1) has shown that speakers reduce temporal overlap in precisely those contexts *where C1 is most at risk of being masked*. These findings have led, logically, to the hypothesis that speakers may control the degree of overlap between consonantal gestures on the basis of knowledge about the perceptibility of the individual gestures. We also consider the acoustic consequences of overlap patterns, in light of the discussions on cue robustness and cue precision in Henke et al. (2012), Wright (1996, 2001), and Benki (2003). Temporal overlap can affect cue precision in particular, defined by Henke et al. as the degree to which information in the acoustic signal narrows down, perceptually, the number of segmental choices available to the listener.

In light of the evidence from production that speakers adjust timing in ways that may enhance recoverability, we aim to verify the basic premise of the C1C2 recoverability hypothesis: Does reduced overlap (longer lag) between consonant gestures help listeners to accurately recover C1 in a C1C2 stop sequence? To investigate this, we focus on the stop-stop sequences in Georgian, a language which has received a great deal of attention for its rich phonotactic combinations that present almost no gaps. The production patterns of Georgian stop-stop sequences have been well studied (Zhgenti, 1956; Chitoran et al., 2002; Crouch, 2022; Crouch et al., 2023a; Pouplier et al., 2022). In Section 2, we review production patterns established for Georgian and for other languages that have served as arguments for or against a perceptual motivation. In Section 3, we present an overview of the relevant perception studies that can inform our question. Section 4 outlines our perception experiments, followed by details of the materials and methods in Section 5 and the presentation of the results in Section 6. We end with a discussion of the results and conclusions in Section 7.

## 2. Perceptual motivations for observed production patterns

In many languages, stop-stop clusters have been found to exhibit variable overlap depending on two factors: (a) position in the word or utterance: initial C1C2 is generally less overlapped than medial; (b) order of place of articulation in the stop sequence: C1C2 with a back-to-front place order is generally less overlapped than front-to-back.

Many studies, reviewed below, have interpreted these overlap patterns based on perceptual recoverability, without testing perception but following a logical reasoning. In utterance-initial

position, for example, the only acoustic information available for C1 is its release burst. In the case of increased temporal overlap, the absence of both the C1 acoustic release and of V-C1 or C1-V transitions makes it difficult to recover C1. Moreover, the recovery of C1 may be more critical word-initially, because word-initial onsets are more important in word recognition than word-medial ones (Marslen-Wilson & Zwitserlood, 1989). Word-medially, by contrast, C1 may still be recoverable from transitions out of a preceding vowel, even at a greater degree of overlap.

A similar argument holds for a back-to-front order of C1C2 stop sequences (e.g., *gd*, *tp*), where an anterior C2 closure will mask a posterior C1 release unless there is an ample temporal distance between the two. If the C2 constriction is anterior in the vocal tract relative to C1, the C1 release may be completely hidden acoustically, unless the C2 constriction occurs after C1 has already been released. But if the C2 constriction is posterior to C1 (e.g., front-to-back sequences such as *dg*, *pt*), some acoustic information will still be present at C1 release, even with substantial overlap.

Evidence for longer lag (reduced overlap) produced word-initially and in back-to-front stop sequences is quite consistent across languages. In Georgian stop-stop sequences, Chitoran et al. (2002) and Crouch (2022) found that consonant timing varies systematically with position in the word and with the place order of the stops in native speakers' productions. Word-initial stop-stop sequences have significantly longer lag than word-internal ones, and sequences with a back-to-front (B-F) order of constriction location (e.g., *gd*, *tp*) have longer lag than sequences with a front-to-back order (F-B) (e.g., *dg*, *pt*). Chitoran et al. (2002) attributed these patterns to considerations of perceptual recoverability, that is, when a stop-stop sequence (C1C2) is produced with short lag, the acoustic release of C1 can be masked by C2 closure.

Beyond Georgian, a longer lag pattern word-initially has been reported for several other languages: English (Hardcastle, 1985, for stop-liquid sequences; Byrd, 1996, for /s/-stop), Tsou (Wright, 1996), Russian (Kochetov & Goldstein, 2005), Moroccan Arabic (Gafos et al., 2010), Hebrew (Yanagawa, 2006). Evidence for a place order effect has been found in English (Hardcastle & Roach, 1979; Byrd, 1992, 1996; Zsiga, 1994; Surprenant & Goldstein, 1998), Tsou (Wright, 1996), Taiwanese (Peng, 1996), Russian (Zsiga, 2000), Korean (Kochetov et al., 2007), French (Kühnert et al., 2006).

As promising as these cross-linguistic production patterns may be for a recoverability hypothesis, it is important to consider several caveats. First, across many of these studies, lag or overlap is quantified by a variety of measures, acoustic and/or articulatory. Also, it is not entirely clear how well the acoustic measures employed (e.g., the acoustic duration of the inter-burst interval) correspond to articulatory measures. Pouplier et al. (2017) further warn that hypotheses about the perceptibility of consonant sequences cannot be tested by measuring articulatory lag alone, since it overlooks important differences between acoustic and articulatory coarticulation effects. For stops in particular, the relation between their articulatory and acoustic release is

dependent on their constriction location in the vocal tract. Articulatory releases, as measured based on kinematic data, may precede acoustic releases to varying degrees. Further, differences in intraoral pressure dynamics affect the resulting acoustics depending on the constriction locations and voicing of the stops involved, and thus can affect perception independently of lag. Pouplier et al. test the cue robustness hypothesis of Henke et al. (2012) and Wright (1996), proposing that typologically rare clusters are perceptually suboptimal. These rare clusters are predicted to have a limited amount of overlap, and this limited overlap is, moreover, predicted to be resistant to coarticulatory changes normally triggered under rate pressure. It is understood that this coarticulatory stability is under speaker control, reflecting speakers' knowledge of perceptual cues to the identity of the segments involved. Nevertheless, Pouplier and colleagues do not find support for the covariation between perceptual recoverability and degree of overlap. Contrary to the predictions, speech rate manipulation did affect degree of overlap regardless of their optimal or sub-optimal phonotactics. At fast speech rate, overlap increased regardless of phonotactic status. They found, instead, a lexical frequency effect, whereby degree of overlap increased significantly more at fast speech for clusters with higher frequency, but they found no support for the auditory cue robustness hypothesis.

Differences between initial and medial lags, for example, have also been explained by prosodic factors, as temporal patterns are known to be influenced by word or prosodic boundaries (Edwards et al., 1991; Sotiropoulou & Gafos, 2022). The boundary as a prosodic event affects the constriction formation and release of consonantal gestures, as well as their acoustic duration, and distance from the boundary is known to modulate the strength of such effects. In AP, these durational patterns are conceptualized as effects of a prosodic time variable, which slows down articulatory events close to prosodic boundaries (Byrd & Saltzman, 2003; Krivokapić, 2007; Cho et al., 2014; Katsika, 2016).

Evidence against the recoverability hypothesis comes from Korean and Japanese, where stop-stop sequences show reversed place order effects. In Korean (Son, 2008) and Japanese (Yanagawa, 2003), front-to-back /pt/ and /pk/ are found to be less overlapped than /tp/ and /kp/, contrary to expectations based on recoverability. Other studies revealed place order asymmetries in contexts where they are not predicted by recoverability, thus weakening the perceptual recoverability hypothesis, as well. For example, Chitoran and Goldstein (2006) found a place order effect in Georgian not only in stop-stop but also in stop-liquid sequences, where it would not necessarily be expected, since in this context C1 is not perceptually vulnerable. Longer lag was found in the front-to-back /kl/ and /kr/ sequences than in /pl/ and /pr/, even though, in both cases, the identity of C1 is not as easily masked at shorter lags as it is in stop-stop sequences. When followed by a liquid, a C1 stop preserves some acoustic information in the transition into C2. A similar, perceptually unmotivated place order effect was found in French stop-/l/ sequences by Kühnert et al., (2006). This longer lag in stop-liquid sequences has instead

been attributed to articulatory characteristics of the liquid, such as stiffness (Du & Gafos, 2023). Assuming that speakers control the stiffness parameter, as proposed in AP, the findings on stop-liquid sequences point to articulatory constraints, independently of perceptual considerations, that can shape timing patterns of adjacent gestures.

Similar place order effects unmotivated by perceptual recoverability have also been observed by Yip (2013) in Greek, a language whose CC inventory approaches that of Georgian. Greek C1C2 sequences show a place order effect at least for some speakers, but in addition to stop-stop sequences, the effect is observed even when one of the consonants is a fricative or when C2 is a liquid. According to recoverability predictions, this result is unexpected, since a fricative in C1 position is not as vulnerable at high overlap as a stop. The frication noise of the fricative carries sufficient acoustic information that is not easily masked by increased overlap with C2. Thus, the Greek and the Georgian production patterns fail to provide specific, conclusive evidence supporting perceptual recoverability, which could be due to an alternative motivation, making it essential to verify the hypothesis directly from the perception side.

### 3. Evidence from perception

Relatively few studies so far have tested the perceptual effect of lag duration and/or of its direct acoustic consequences (i.e., the presence vs. absence of an acoustic release burst or vocalic transition) on the perceptual recovery of C1. We review these perceptual studies in this section.

Surprenant and Goldstein (1998) studied the effect of gestural overlap on the perception of the American-English stop sequences /t#p/ and /p#t/ across a word boundary, using stimuli extracted from x-ray microbeam data. The study addresses the place order effect. In one experiment, listeners were asked to perform a consonant monitoring task for stimuli containing stop-stop sequences (*tot#puddles*, *top#tuddles*) and controls containing C1 single stops (*tot#huddles*, *top#huddles*). Results showed that: (i) C1 in the stop-stop sequences was detected significantly less often than the C1 single stops, (ii) a C2 bilabial gesture more often obscured a preceding C1 alveolar gesture than the reverse order, and (iii) the detection rate correlated with the degree of overlap between lip and tongue tip gestures. In a second experiment, the acoustic burst—acoustic information crucial to retrieve lip and tongue tip gestures at the articulatory release—was removed from the stimuli. The absence of the acoustic burst decreased the detectability of the stops generally, but the asymmetry between the detectability of C1 in /t#p/ vs. /p#t/ did not entirely disappear. This specific result is significant, because it suggests that the timing of the two gestures alone may have perceptual consequences independently of the acoustic consequences of that timing.

Their study, using natural speech data, confirmed the results of an earlier perception study by Byrd (1992), which used synthetic stimuli obtained with the Haskins articulatory synthesizer (Rubin et al., 1981). Byrd varied the amount of overlap in *bab#dan* and *bad#ban*. Results of

a forced choice identification test showed that as overlap increased, C1 identification was significantly reduced, and C1 was perceived as assimilated to C2. Also, consistent with the place order effect, the alveolar C1 in back-to-front /d#b/ was affected at a much smaller degree of overlap than the labial C1 front-to-back in /b#d/.

Coronal-labial and labial-coronal sequences were also studied by Chen (2003) using computational modelling to test the effects of gestural overlap on the recoverability of C1. The stimuli were generated with GEST (Browman & Goldstein, 1990; Gafos, 2002), with acoustics derived with the Haskins articulatory synthesizer. Results based on the listener model found that increasing overlap in labial-coronal sequences had little effect on C1 recovery, but in the opposite coronal-labial order, recoverability rates for C1 decreased.

These studies are limited to combinations of labial and coronal stops across a word boundary, a context where intergestural timing is known to be highly variable. Directly relevant to the current study are the studies on segmental recoverability as a function of lag duration within a syllable, and of the presence vs. absence of a stop release burst. Both were investigated by Wright (1996) in Tsou word-initial clusters. One experiment manipulated the acoustic release burst of C1: The burst was removed from a word-initial cluster (/tmihi/ 'to hang') and inserted before a singleton C (/mihi/ 'to desire a denied thing'). When the release burst was added to a singleton, Tsou native listeners reported hearing clusters. When the release burst was removed from a cluster, they reported hearing the singleton C. A second experiment varied the acoustic inter-burst interval (IBI), representing the acoustic lag between C1 and C2 in the initial stop-stop cluster /pt/. Listeners' perception varied with the acoustic lag duration. When the C1 and C2 bursts were separated by 0 – 25 ms silence, listeners reported hearing a single onset /t/, while between 50 and 150 ms, they reported hearing the full cluster. The cluster response went down again beyond 150 ms, and the single /t/ response went up instead.

The perceptual role of acoustic details in stop-stop sequences was later reliably established in Wilson et al., (2014), in a cross-language perception study. While the study did not consider lag/overlap directly, it showed that masking acoustic details of C1 in a stop-stop sequence can affect its recoverability, at least for non-native listeners. The authors manipulated the amplitude and duration of C1 release bursts in initial stop-stop sequences. The stimuli were non-words containing stop-stop sequences licit in Russian and illicit in English, produced by a Russian-English bilingual. Native English speakers were then asked to hear and produce the sequences in a shadowing task. The results showed that greater burst amplitude made C1 more likely to be correctly produced, protecting it from deletion or other modifications. When C1 burst amplitude decreased, C1 underwent significantly more deletion and change.

The presence vs. absence of stop releases was also tested in a cross-linguistic perceptual study by Kochetov and So (2007), who conducted two experiments using Russian voiceless clusters with released and unreleased stops. Results from native listeners of Russian, Canadian English,

Korean, and Taiwanese Mandarin showed that the presence or absence of stop releases strongly affected the perceptual accuracy of place of articulation.

All these previous findings collectively suggest that a stop release is crucial in perceiving the stop. When the release is masked or removed, the consonant becomes less likely to be accurately perceived. Therefore, the temporal pattern that is likely to mask the release of C1 is predicted to be avoided in production (as supported by many production studies), and listeners are less likely to perceive the intended consonant when the lag is short (or overlap is high), thus the release is less (or not) audible.

In addition to differences in lag and release bursts, previous studies have also reported the variable production of vocalic releases of C1 in a stop-stop sequence in several languages (for a recent review, see Hall, 2024). These vocalic releases, or vocoids, are typically seen as the acoustic consequence of C-C coordination with longer lag (reduced overlap), and have been reported for Moroccan Arabic (Gafos, 2002), Tashlhiyt (Ridouane & Fougeron, 2011; Ridouane & Cooper-Leavitt, 2019), and Georgian (Chitoran et al., 2002; Goldstein et al., 2007; Crouch et al., 2023a). By providing formant transitions, these transitional vocoids may help convey information about the identity of C1.

Evidence from a perception study on Tashlhiyt word-initial consonant sequences confirms this prediction. Zellou et al. (2024) tested the perception of vowelless Tashlhiyt words across clear and casual speaking styles by native and non-native naïve listeners. A part of this study examined the role of the transitional vocoids, establishing that the longer duration of these vocoids in consonant sequences improves native speakers' discrimination of CCC words. When discrimination was more challenging, such as between pairs with sequences of matching sonority (falling vs. falling, or plateau vs. plateau), the presence of a transitional vocoid improved discrimination for both native and non-native listeners. A longer duration of the vocoid was also beneficial. CCC words were easier to discriminate when they contained longer transitional vocoids, and this effect was particularly strong for word pairs with sonority plateaus.

The Tashlhiyt results may inform predictions about the perception of Georgian clusters, although the two languages are not directly comparable. While in Georgian all complex word onsets constitute one syllable onset, in Tashlhiyt they can be heterosyllabic, since any consonant can be syllabic (Dell & Elmedlaoui, 2002; Ridouane, 2016; Ridouane et al., 2014). While a comparison of Georgian and Tashlhiyt lies beyond the scope of this study, it is important to note that the two languages differ in the distribution of the vocoids. In both languages, the distribution of vocoids is sensitive to the sonority sequencing, but they show different distributional patterns. The Georgian speakers' production data (Crouch et al., 2023a; Crouch et al., 2023b) revealed that vocoids appear predominantly in sonority rises (56% of the data), less often in sonority plateaus (25% of the data), and only rarely in sonority reversals (9% of the data). The situation

seems to be the opposite in Tashlhiyt (Zellou et al., 2024). These differences suggest that the transitional vocoids may not necessarily result from similar gestural coordination patterns in the two languages. In Georgian, vocoids occur between the two consonants, and they can be attributed to an overall reduced C-C overlap pattern within a complex onset (supported by results of a seven-language comparison in Pouplier et al., 2022). In Tashlhiyt, however, the vocoids may appear in various positions, either before the consonant sequence, or breaking it. Since Tashlhiyt consonant sequences, unlike those of Georgian, are heterosyllabic, they may involve variable modes of coordination. This difference in coordination may in turn influence how consonant sequences are perceived by listeners.

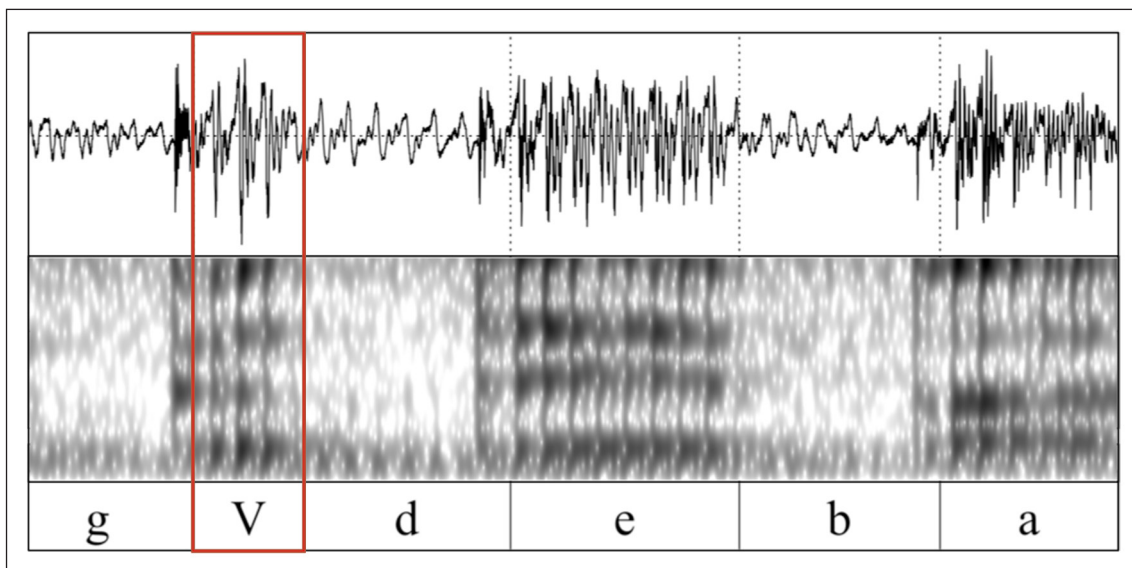
A previous perception study conducted with Georgian listeners suggests that the listeners' perception is indeed influenced by temporal organization. Kwon and Chitoran (2024) tested Georgian listeners on C<sub>Ca</sub>-CVC<sub>a</sub> discrimination in French stimuli. The results revealed perceptual confusion in particular for the C<sub>Ca</sub>-C<sub>ø</sub>C<sub>a</sub> contrast, where the French vowel /ø/ is phonetically similar to the C-C transition in Georgian, both in terms of temporal organization and tongue shape (formant structure). This suggests that Georgian listeners use their native knowledge of the temporal implementation of word-onset CC sequences in responding to the task. The present study will provide new information that can be connected to these earlier results, aiming to understand the relationship between the vocalic transitions and timing lag. Ultimately, it will help verify what listeners use—vocalic transitions, or lag alone—and will further help us understand what speakers plan.

With these considerations in mind, we conducted a perception study using Georgian as a test language. The relatively unconstrained phonotactics of the language allow us to test perception of naturally produced C<sub>1</sub>C<sub>2</sub> stop combinations of varying degrees of overlap, utilizing the same sequences produced in word onset and word medial positions. We first establish the perceptibility of C<sub>1</sub> in such sequences. Then, following Pouplier et al.'s (2017) cautionary remark against relying exclusively on articulatory lag for perceptibility hypotheses, we test whether, and to what extent, perception accuracy can be related to measures of temporal overlap (lag) between the two consonantal gestures, as measured from an articulatory signal, and to its acoustic consequences. To the best of our knowledge, our study is the first to test the relevance of both articulatory and acoustic patterns for recovering phonological information about C<sub>1</sub> in stop-stop sequences.

#### **4. The current perception experiment**

Chitoran et al. (2002) proposed that the variation present in Georgian production patterns can be interpreted as speaker-controlled strategies for increasing C<sub>1</sub> perceptibility in a stop-stop context, where C<sub>1</sub> gestures are harder to recover. The reasoning behind the interpretation is based on the following key observations:

- The two Georgian speakers whose kinematic data were analyzed in their study consistently produced a longer lag in contexts where C1 could be easily masked. It was proposed that the longer lag would prevent C1 release from being masked by C2 closure and would allow for a clearer, audible C1 release in a stop sequence;
- The speakers occasionally produced a C1 vocalic release, resulting in a transitional vocoid, as illustrated in **Figure 1**. Vocoids were observed more often in voiced stop-stop sequences and in sequences with a back-to-front place order, where C1 release is less likely to be audible because of the following anterior constriction. It was thus proposed that the vocoid would provide clearer C1 formant transitions, which could contribute to the accurate recovery of C1.



**Figure 1:** Example of a vocalic release (transitional vocoid) occurring between C1 and C2, in the word /dgeba/ ‘to be thrown’.

#### 4.1 Hypotheses and predictions

Based on the interpretation of the observed Georgian production patterns, we test the following hypotheses:

**H1:** Longer timing lag (or reduced overlap) between C1 and C2 facilitates recovery of C1 gestures

**H2:** A C1 vocalic release facilitates recovery of C1 gestures

In the present study, we test the hypotheses in two perception experiments with native listeners of Georgian. The perceptual responses are then evaluated against detailed acoustic and articulatory (EMA) analyses of the Georgian stimuli.

Experiment 1 is a forced choice identification task, in which listeners heard CCV portions excised from words containing stop-stop sequences. They were asked to decide whether the short sound they heard began with a CV or a CC sequence. It is predicted that listeners would more accurately identify stimuli as beginning with CC when timing lag is longer, and thus perceptual recoverability is assumed to be enhanced. The best CC identification rates are therefore expected for target stimuli based on initial back-to-front sequences (e.g., #gdV), which are the least overlapped (have the longest lag), and the poorest rates are expected for stimuli containing word-medial front-to-back sequences (e.g., dgV), which are the most overlapped (have the shortest lag).

Experiment 2 was conducted a week later, with the same participants. It consisted of a transcription task. Each listener was exposed only to those stimuli for which they previously gave a CV answer and asked to transcribe the short sequence by hand. This transcription task determined whether listeners responded CV in Experiment 1 because they failed to detect C1 in the stop-stop sequence, or because they detected a vocalic portion between C1 and C2, and treated it as a vowel. Experiment 2 thus complements Experiment 1 by allowing for reliable answers to each of the two questions of the study:

- (a) Do listeners ever miss C1?
- (b) Do listeners detect a C1 vocalic release when present, and does it help them to accurately identify C1?

## 5. Material and methods

### 5.1 Acoustic and articulatory data collection

The stimuli were prepared based on the acoustic and articulatory Georgian data previously collected at Haskins Laboratories, in New Haven, CT, and analysed in Chitoran et al., (2002). One specific speaker was selected to provide the stimuli, based on the fact that he produced more frequent vocalic transitions between C1 and C2 in his speech. The speaker, a male in his mid 20s, had been living in the United States for approximately two years at the time of the recording and reported using Georgian on a regular daily basis. He reported no speech or hearing impairment. Prior to the start of the experiment, the speaker was informed of the purpose of the study; he read and signed consent forms (all of this was done in English).

Simultaneous acoustic and kinematic data were collected for stimuli, including stop-stop sequences and filler items (**Table 1**). Target stimuli were recorded along with other distractors, which included CC combinations other than stop-stop. The stop-stop sequences in this study are identical tokens to the ones whose production was analyzed in Chitoran et al. (2002). Each target word was produced in the carrier phrase [sit'q'wa \_\_\_ gamoit<sup>h</sup>k<sup>h</sup>mis orɔʒer] ('The word \_\_\_ is pronounced twice'). A computer screen presented the sentences one at a time, in Georgian script. The speaker was invited to read each sentence aloud at a normal pace. If he paused or had

a false start, he was asked to re-read the entire sentence. Fourteen repetitions of each sentence containing a stop-stop target word were presented in randomized order and recorded.

	Front-to-back			Back-to-front		
	stimuli	words used		stimuli	words used	
<b>Word-initial</b>	bge	bgera	'sound'	gbe	g-ber-av-s	'is inflating you'
	p <sup>h</sup> t <sup>h</sup> i	p <sup>h</sup> t <sup>h</sup> ila	'hair lock'	t <sup>h</sup> be	t <sup>h</sup> b-eb-a	'it is warming up'
	dge	dg-eb-a	's/he stands up'	gde	gd-eb-a	'to be thrown'
<b>Word-medial</b>	bga	abga	'saddle'	gbe	da-gbera	'to say the sounds'
	p <sup>h</sup> t <sup>h</sup> a	ap <sup>h</sup> t <sup>h</sup> ar-i	'hyena'	t <sup>h</sup> ba	ga-t <sup>h</sup> b-a	'it has become warm'
	dge	a-dg-eb-a	's/he will stand up'	gde	a-gd-eb-a	'throw in the air'
<b>Fillers</b>	t'k'e	t'k'ena	'to hurt'	k'bi	k'bili	'tooth'
	t'k'a	bat'k'an-i	'lamb'	k <sup>h</sup> t'i	k <sup>h</sup> t'it <sup>h</sup> or-i	'founder'
	bra	braz-i	'anger'	k're	k'reba	'meeting'
	t <sup>h</sup> k <sup>h</sup> e	albat <sup>h</sup> #k <sup>h</sup> er-i	'probably barley'	k <sup>h</sup> t'e	p <sup>h</sup> ak <sup>h</sup> t'-eb-i	'facts'
				t <sup>h</sup> ba	albat <sup>h</sup> #bay-ji	'probably in the garden'
<b>CVCV controls</b>	bile, t'ebi, deba, geba					

**Table 1:** Words used to create perception stimuli ('-' indicates a morpheme boundary).

Intervocalic multiple consonant sequences in Georgian are reported to be syllabified as a simplex coda followed by a complex onset (Harris, 2002). However, when the sequence consists of only two consonants, the syllabification intuitions of native speakers vary, oscillating between a complex onset (V.CCV) and a coda + onset (VC.CV).

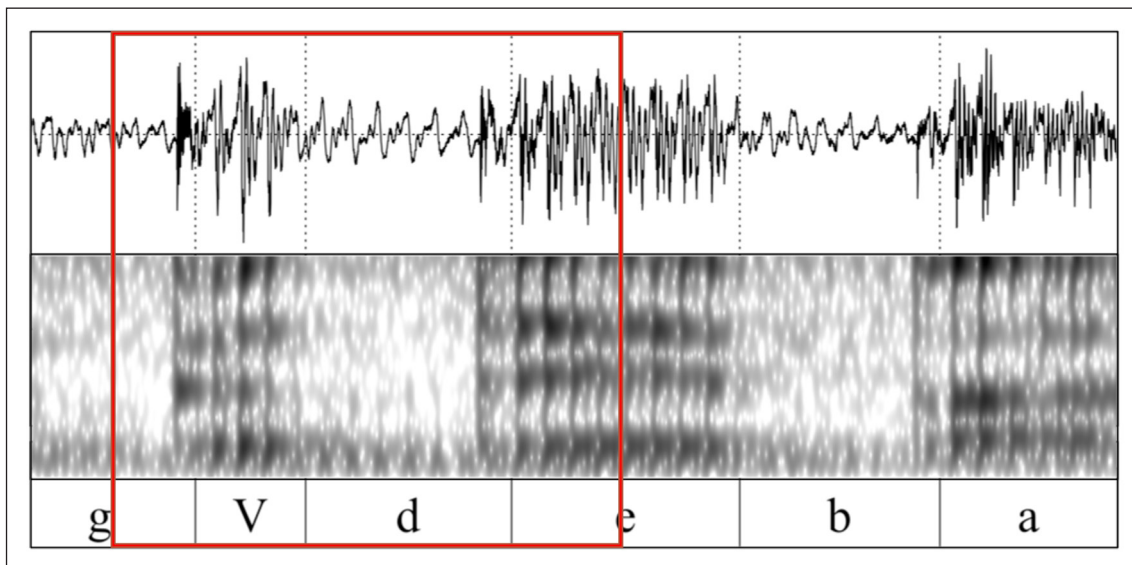
The EMA magnetometer system (Perkell et al., 1992), in use at Haskins at the time, was used for data collection. Two receiver coils were attached to two midsagittal points on the tongue: approximately 1 cm from the tongue tip (TT), and tongue dorsum (TD) as far back as possible,

to capture velar constrictions. One coil each was placed on the upper and lower lip on the midsagittal plane. Reference coils were placed on the upper and lower teeth and on the nose bridge, and were used for head movement correction.

## 5.2 Stimuli preparation

From the recorded C1C2 sequences, six different stop-stop sequences – /bg, dg, pt, gb, gd, tb/ – varying in their place order (three front to back F-B, three back to front B-F) were selected as the target stimuli. Other C1C2 sequences that are not the target six sequences were used as fillers, in order to vary the types of stimuli to which the listeners were exposed. Multiple productions (three to five) were included for each sequence in order to examine the effect of timing variation naturally present in Georgian CC production. For this speaker, one /gat<sup>h</sup>ba/ and two /t<sup>h</sup>beba/ tokens were lost due to mispronunciation.

Some of the selected productions included C1 vocalic releases, which we indicate henceforth with a superscript <sup>v</sup>: C1<sup>v</sup>C2. The selected stimuli were excised C1C2V portions (**Figure 2**), segmented on the acoustic signal from the midpoint of C1 closure to the midpoint of V, adjusted to the nearest zero-crossing, and avoiding coarticulation with the following consonant.



**Figure 2:** Example of segmentation for the stimulus /g<sup>v</sup>de/ extracted from the word /gdeba/.

Four C1VC2V sequences were included as controls: [deba], [geba], [bile], [t'ebi]. These were the last two syllables excised from the trisyllabic words [agdeba], [adgeba], [sat'k'bile], [p<sup>h</sup>ak<sup>h</sup>t'ebi], from the midpoint of the C1 closure to the midpoint of the final vowel. In these

four Georgian words, the initial syllable is stressed. By removing it, we made sure there was no prominence on the first vowel in the remaining CVCV sequences. All target stimuli and controls were attested word onsets in Georgian. They were segmented and normalized for intensity using Praat (Boersma & Weenink, 2021).

### 5.3 Participants and procedure

The perception experiment was conducted in Tbilisi, Georgia, in a quiet office in the Linguistics Department at Tbilisi State University. Twenty-nine Georgian native listeners (17 female) participated in two experiments. A Georgian student research assistant was hired to recruit participants; these were students at Tbilisi State University, recruited via fliers and the assistant's personal contacts. The assistant explained to the participants, in Georgian, the information letter, the consent forms, and the experiment instructions.

In Experiment 1, the forced choice identification task, a total of 222 stimuli were presented to, and randomized across, each participant. These included 114 target stimuli, 70 fillers, and 38 CVCV controls. After hearing each stimulus, the listeners were asked to identify whether it begins with a sequence of two consonants 'cc', or with a consonant-vowel sequence 'cv'. Before the experiment, each participant confirmed that they were familiar with the terms 'consonant' and 'vowel', and were able to give accurate examples of each.

Experiment 2 was conducted a week later with the same participants, to disambiguate the 'cv' response in Experiment 1. In Experiment 1, listeners could have responded 'cv' for a stimulus such as /bge/, for example, if they heard a vowel between the stops [b<sup>v</sup>ge], but also if they heard only one of the stops [ \_ge] or [b\_e]. Experiment 2 thus consisted of a transcription test, in which listeners heard subsets of the stimuli from Experiment 1. Each participant heard all tokens (three to five) of any stimulus item for which they had given at least one 'cv' response in Experiment 1, and was asked to transcribe what they heard by hand in Georgian orthography. This ensured that multiple productions of the same word were included when applicable. Georgian orthography being phonemic, IPA transcription of the responses is relatively easy. It was done by the first author, who is familiar with the Georgian alphabet. The native Georgian research assistant also verified the participants' transcriptions.

Both experiments were conducted on a Windows laptop computer, using a program written and kindly provided by René Carré and Emmanuel Ferragne. Both experiments were self-paced, and each one lasted between 20 and 25 minutes. For Experiment 1, the forced choice identification task, participants were seated in front of the computer, and the stimuli were played back via headphones, one at a time. They were asked to listen to each sound and respond by pressing either the F or the J key on the keyboard. The F key was labelled as 'cv', and the J key as 'cc', using Georgian letters: თხ [t<sup>h</sup>x] for 'cv', თანხმოვანი – ხმოვანი [t<sup>h</sup>anxmovani–xmovani]

'consonant–vowel', and თთ [t<sup>h</sup>t<sup>h</sup>] for 'cc', თანხმოვანი – თანხმოვანი [t<sup>h</sup>anxmovani–t<sup>h</sup>anxmovani] 'consonant–consonant'. After each response, the participants clicked on the screen to move on to the next stimulus.

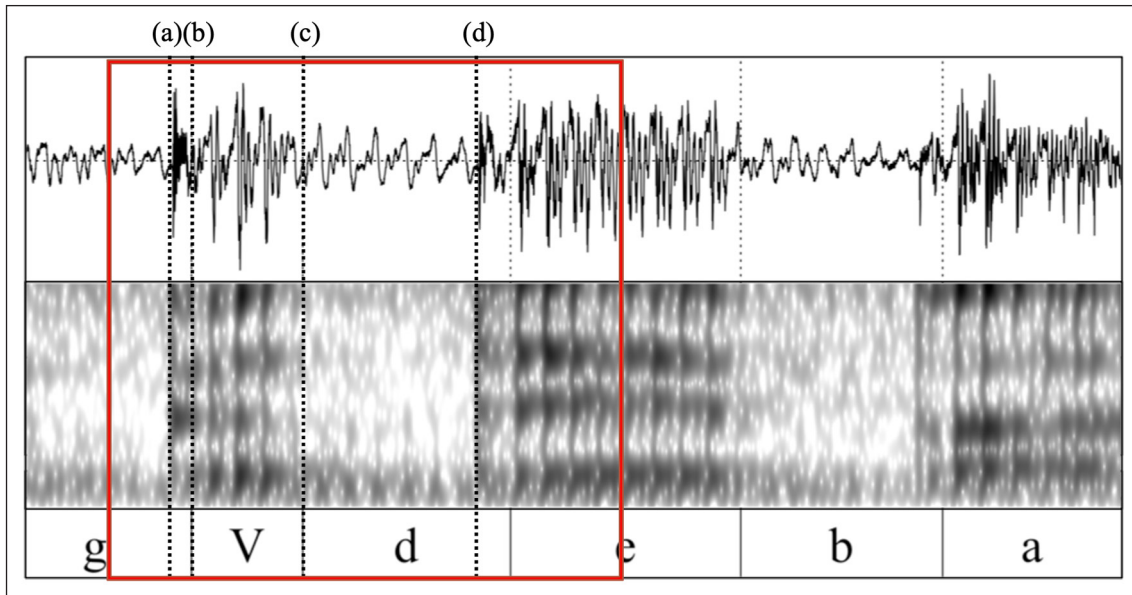
Experiment 1 was preceded by two practice blocks. In the first block, participants were asked to simply listen to 10 stimuli, to familiarize themselves with the short sounds they would hear. In the second practice block, they listened and responded to 10 different stimuli. This was to make the participants familiar with the task, and thus no feedback was provided for the practice trials. The 20 practice stimuli were not included in the actual experiments.

We would like to acknowledge that using perception stimuli recorded with EMA might raise concerns about possible speech distortions and intelligibility due to the EMA sensors. To the best of our knowledge, the effects of EMA sensors on the perception of speech are largely unexplored. Several studies have tested the effect of sensors on production, comparing typical and disordered speech such as aphasia, apraxia, and Parkinson's Disease (e.g., Katz et al., 2006; Tienkamp et al., 2024). These studies reported some interference in the production of sibilant fricatives and in the acoustic-articulatory vowel space. While we cannot rule out the possibility that the presence of EMA sensors may have interfered to some extent with the Georgian speaker's production, our stimuli come from a single speaker and were carefully selected based on the articulatory measure of overlap, on which our hypotheses are crucially based. Since testing an articulatory hypothesis is the main goal of our study, we prioritized this latter point. We were ultimately reassured by the relatively high accuracy scores of the transcriptions obtained in Experiment 2: C1 was correctly transcribed 66% of the time and C2, 75% of the time. These scores do not reflect serious intelligibility issues. Moreover, the higher accuracy of C2 transcriptions is what we expected to find, given that a vowel always followed C2. This suggests that sensor interference, if any, was limited.

#### 5.4 Acoustic and articulatory analysis of the stimuli

To understand the relation between the participants' responses and the phonetic properties of the stimuli, we analyzed the acoustic and articulatory properties related to the timing lag (or overlap) between C1 and C2, as well as those related to the vocalic release of the C1. The following three acoustic parameters were measured:

- 1) *Acoustic Lag*: the duration of the inter-burst interval, measured from C1 release burst (**Figure 3a**) to C2 release burst (**Figure 3d**);
- 2) *Presence of Vocalic Release*: the occurrence of vocalic releases (present vs. absent);
- 3) *Vocalic Release Duration*: the duration of vocalic releases, when present, measured from **Figure 3b** to **Figure 3c**.



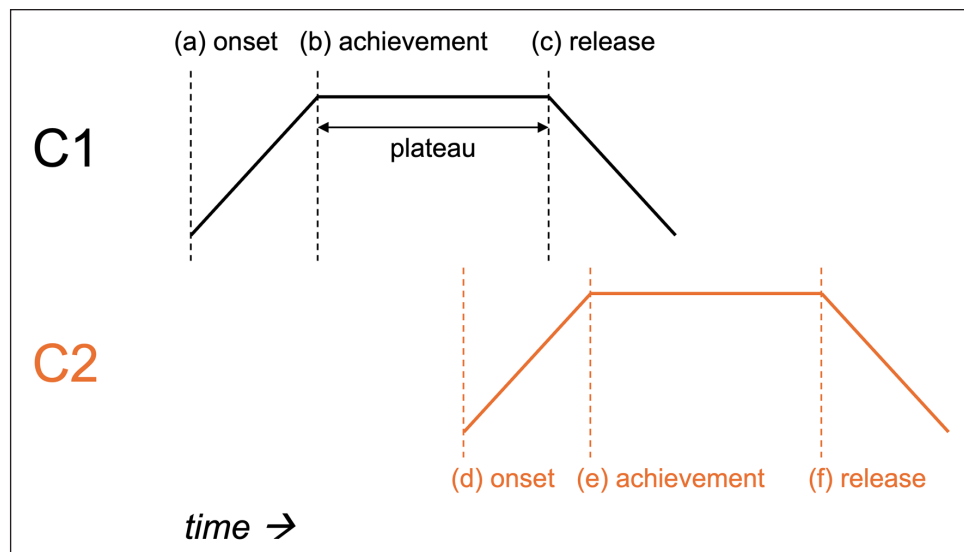
**Figure 3:** Acoustic landmarks used in measurements: (a) C1 release burst; (b) vocalic release onset (if present); (c) vocalic release offset (if present); (d) C2 release burst.

Three articulatory measures were examined, based on the articulatory landmarks measured on the EMA signal:

- 1) *Release Lag*: Temporal distance from C1 release (4c) to C2 release (4f);
- 2) *Onset Lag*: Temporal distance from C1 gesture onset (4a) to C2 gesture onset (4d);
- 3) *Relative Overlap*:  $1 - [C2 \text{ gesture onset (4d)} - C1 \text{ achievement (4b)}] / [C1 \text{ release (4c)} - C1 \text{ achievement (4b)}]$  (Gafos et al., 2010; Roon et al., 2021)

In EMA signals, movement trajectories of the receiver coils attached to the tongue tip, tongue dorsum, upper lip, and lower lip were evaluated. The articulatory constriction formation and release were identified using the Matlab analysis program, *Mview* (provided by Mark Tiede). The *Mview* algorithms allow the computation of articulatory landmarks based on the velocity profiles of the relevant sensors. The peak velocities of the constriction formation and release movements were calculated algorithmically. For each gesture, the following three points were identified and labelled, using a 20% threshold of the velocity peaks: the gesture onset, constriction (target) achievement, and constriction release. For labials, the Euclidean distance between upper and lower lip receiver coils was used to compute lip aperture, and thus measure labial constrictions. For coronals, the gestural landmarks were determined by evaluating the distance of the tongue tip receiver coil to the closest point on the palate. For dorsals, the vertical position of the tongue dorsum coil was used.

The relative overlap measure in (6), like the overlap measure in Chitoran et al. (2002), essentially quantifies what proportion of the C1 plateau (from achievement to release) is free from the influence of the C2 gesture (see **Figure 4**). However, as the relative overlap measure is calculated by subtracting the overlap measure in Chitoran et al. from one, the current measure straightforwardly corresponds to the degree of relative overlap: Greater values thus correspond to the greater degrees of overlap, or the greater influence of C2 movement on C1.



**Figure 4:** Schematic representations of the articulatory landmarks used in measurements.

As expected, some measures are correlated. When all stimuli were considered (**Table 2**), all measures except for the acoustic lag and the release lag showed significant correlations in the expected direction (i.e., positive correlations among the lag measures and negative correlations between the lag measures and the overlap measure). When considering the stimuli that had vocalic releases (**Table 3**), the duration of the vocalic release does not correlate significantly with any of the lag or overlap measure.

	Acoustic Lag	Onset Lag	Release Lag	Relative Overlap
Acoustic Lag	1.000			
Onset Lag	0.417**	1.000		
Release Lag	0.075	0.271*	1.000	
Relative Overlap	-0.297*	-0.781***	-0.450***	1.000

**Table 2:** Correlation matrix for target stimuli measurements.

	Vocalic release	Acoustic Lag	Onset Lag	Release Lag	Relative Overlap
Vocalic release	1.000				
Acoustic Lag	0.189	1.000			
Onset Lag	0.300(*)	0.572***	1.000		
Release Lag	-0.269	0.168	0.150	1.000	
Relative Overlap	0.007	-0.482**	-0.813***	-0.314(*)	1.000

**Table 3:** Correlation matrix for the subset that included vocalic releases.

We also tested whether the presence of a vocalic release was associated with gestural timing and overlap measures using point-biserial correlations. None of the correlations reached statistical significance: acoustic lag [ $r = .20$ ,  $p = .13$ ], onset lag [ $r = .17$ ,  $p = .19$ ], release lag [ $r = -.11$ ,  $p = .41$ ], and relative overlap [ $r = .18$ ,  $p = .18$ ]. These results indicate that the presence or absence of a vocalic release does not reliably co-vary with timing lag or gestural overlap.

## 6. Results

Results of Experiment 1 showed that Georgian listeners successfully identified the presence of a two-consonant sequence in 71% of the C1C2V stimuli. This average is based on 28 out of the 29 participants. One participant was excluded from the analysis because their performance was more than two standard deviations below the group mean (40% correct). In the 29% when the participants responded 'cv', they may have either heard an epenthetic vowel between C1 and C2, or failed to hear either C1 or C2.

Experiment 2 aimed to test whether 'cv' responses reflected misperception of C1 or C2, or the perception of an inserted vowel. C1 was correctly transcribed 66% of the time, including fully correct transcriptions (e.g., /p<sup>h</sup>t<sup>h</sup>a/ → <p<sup>h</sup>t<sup>h</sup>a>), vowel-insertion cases (e.g., /gde/ → <gade>), and cases where C1 was identified correctly but C2 was not (e.g., /t<sup>h</sup>ba/ → <t<sup>h</sup>va>). The remaining 34% included errors such as C1 deletion (e.g., /bga/ → <ga>), C1 laryngeal-category change (e.g., /gbe/ → <k<sup>h</sup>be> or <k<sup>h</sup>obe>), C1 place change (e.g., /p<sup>h</sup>t<sup>h</sup>a/ → <k<sup>h</sup>t<sup>h</sup>a>), C1 manner change (e.g., /bga/ → <vga> or <mga>), metathesis (e.g., /t<sup>h</sup>ba/ → <bt<sup>h</sup>a>), or unrelated responses (e.g., /dge/ → <rio>). For comparison, C2 was correctly transcribed in 75% of trials, consistent with the expectation that C2 benefits from additional cues in the transition to the following vowel. Some C2 errors, such as /b/ transcribed as <v>, may reflect slight articulatory interference associated with the EMA sensors. However, such effects appear limited, as overall C2 accuracy remained relatively high. As the focus of this study is on C1, we do not further analyze C2 here.

We focus on the correct identification of C1, examining how each of the six measures contributes to correct identification of C1 and whether its effect interacts with Place Order (B-F vs. F-B) and C1 voicing (voiceless vs. voiced). As the listeners experienced all stimuli beginning with the C1(V)C2 sequence, we did not consider initial vs. medial word position in these analyses.

### 6.1 H1: Longer timing lag (or decreased overlap) facilitates the recovery of C1

To evaluate H1, we considered four measures: acoustic lag, onset lag, release lag, and relative overlap. They are not independent from one another (see **Tables 2** and **3**), and thus we built a separate series of mixed effect logistic regression models for each measure, using the *lme4* package (Bates et al., 2015) in R (R Core Team 2022). The models predicted the likelihood of correct C1 identification (binary outcome, correct vs. incorrect), based on the fixed effects of Place Order (B-F vs. F-B) and C1 Voicing (voiceless vs. voiced), along with one of the four measures under consideration. For random effects, by-subject and by-item intercepts, as well as all possible random slopes, were considered. We began by building the full model with all possible interaction terms and the fullest random effects structure, and then found the optimal model by removing the interaction terms and the random slopes that do not contribute to the model fit. Nested models were compared using likelihood ratio tests, and AIC/BIC values were examined to confirm consistency across model comparisons. In all comparisons, AIC, BIC, and likelihood ratio results were consistent.

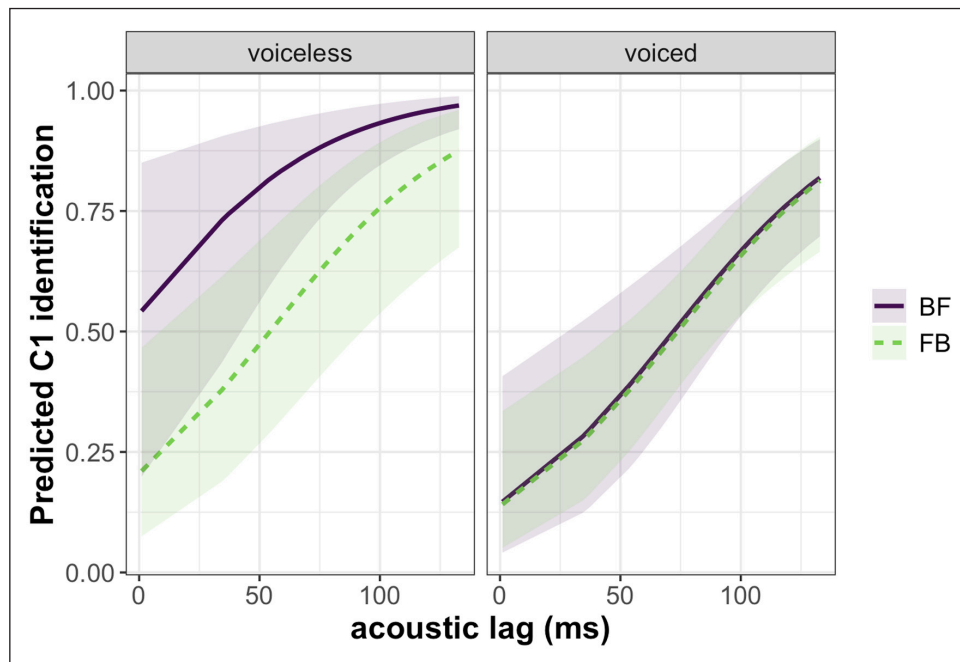
As the timing measures are expected to co-vary with the place order, at least to some extent, we evaluated the multicollinearity with VIF (variation inflation factors) and removed the predictors that are highly correlated with the measure of interest (indicated by VIF values greater than 5, e.g., Shrestha, 2020). This was to build the most reliable model that can best predict the C1 identification using the measure under consideration (i.e., other predictors were less important than the measure of interest). **Table 4** summarizes the best models for each of the four measures examined.

Measure	Best mixed-effect logistic regression model
Acoustic Lag	<b>Acoustic lag</b> + Place Order * C1 Voicing + (1 + Place Order + C1 Voicing  Subject) + (1  Item)
Onset Lag	<b>Onset lag</b> + Place Order * C1 Voicing + (1 + Place Order + Onset Lag + C1 Voicing  Subject) + (1  Item)
Release Lag	<b>Release lag</b> * Place Order + Place Order * C1 Voicing + (1 + Place Order + Release Lag + C1 Voicing  Subject) + (1  Item)
Relative Overlap	<b>Relative Overlap</b> + Place Order * C1 Voicing + (1 + Place Order + Relative Overlap + C1 Voicing  Subject) + (1  Item)

**Table 4:** Best models for each measure.

To determine whether each of the measures contributes significantly to C1 identification, the best models in **Table 4** were compared with the model without the measure in question, using the likelihood ratio tests. If inclusion of the measure improves the model fit, we concluded the measure facilitates C1 identification.

Each of the four measures significantly contributed to the model fit, supporting H1. First, C1 identification was influenced by *Acoustic Lag* [ $\chi^2(1) = 15.02, p = 0.0001^{***}$ ], *Onset Lag* [ $\chi^2(1) = 15.30, p < 0.0001^{***}$ ], and *Relative Overlap* [ $\chi^2(1) = 20.04, p = 0.0001^{***}$ ]. The likelihood of correct C1 identification increased, as *Acoustic Lag* [ $\beta = 0.025, z = 3.91, p < 0.0001^{***}$ ] and *Onset Lag* [ $\beta = 0.027, z = 3.97, p < 0.0001^{***}$ ] increased and *Relative Overlap* [ $\beta = -0.393, z = 4.70, p < 0.0001^{***}$ ] decreased, as expected. As shown in **figures 5, 6, and 7**, these effects were consistent across Place Order or C1 Voicing.



**Figure 5:** Predicted C1 ID accuracy by acoustic lag (ms).

On the other hand, the best model for *Release Lag* included a significant interaction between the measure of interest (release lag) and place order [ $\chi^2(1) = 14.19, p = 0.0002^{***}$ ]. To understand the significant interaction better, post-hoc analyses were conducted using `emmeans()` in the `emmeans` package (Lenth, 2022). The post-hoc analyses revealed that the influence of *Release Lag* differed significantly in back-to-front and front-to-back sequences [ $z = 3.90, p = 0.0001^{***}$ ]: Longer release lags led to a greater chance of correct C1 identification in back-to-front sequences (slope = 0.014, [CI: 0.007, 0.021]), but not in front-to-back sequences (slope = -0.004, [CI: -0.011, 0.003]), as shown in **Figure 8**.

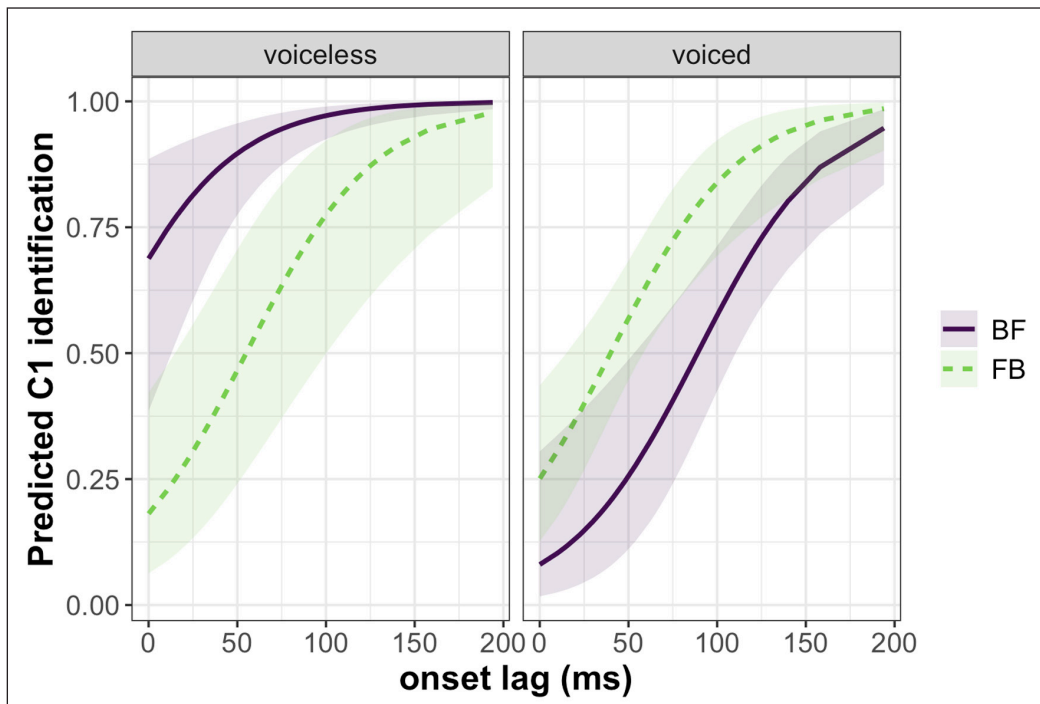


Figure 6: Predicted C1 ID accuracy by onset lag (ms).

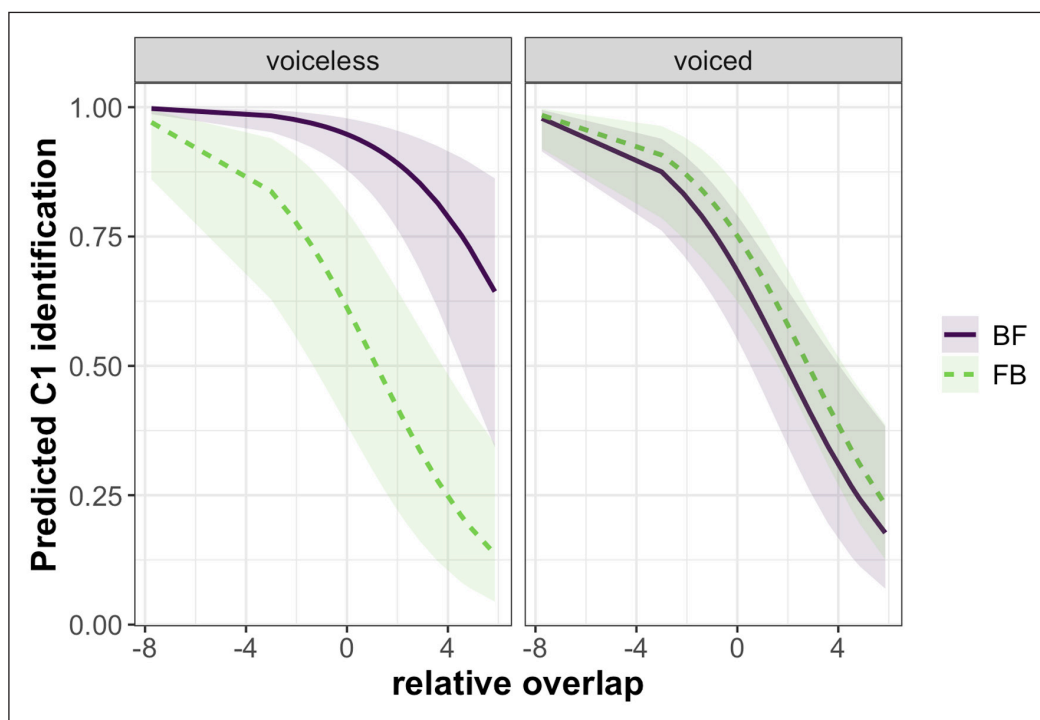


Figure 7: Predicted C1 ID accuracy by relative overlap.

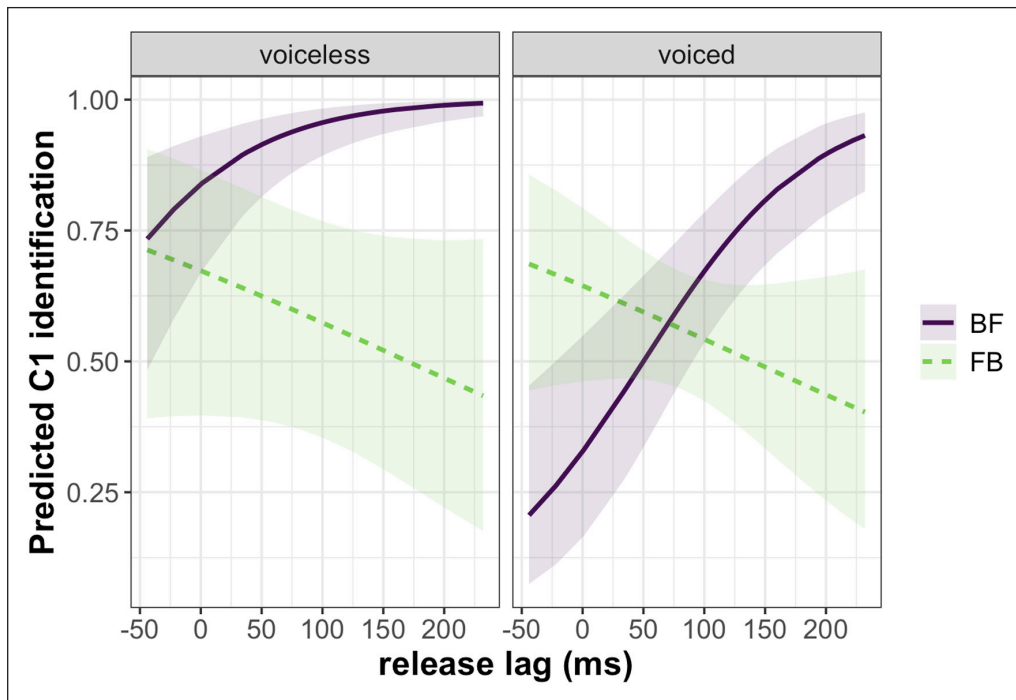


Figure 8: Predicted C1 ID accuracy by release lag (ms).

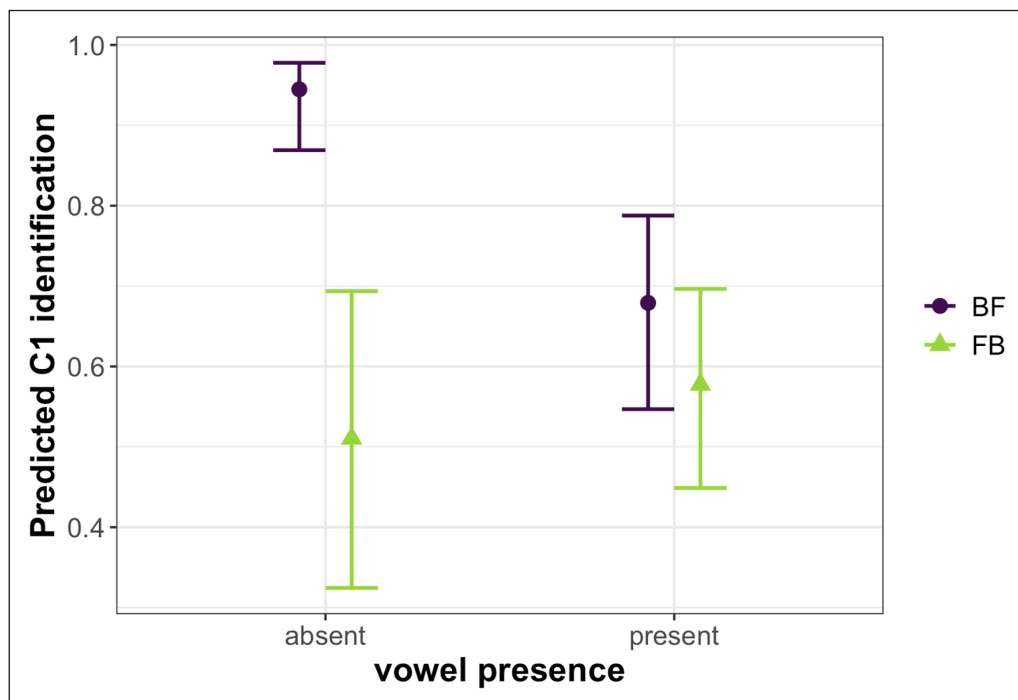
## 6.2 H2: A C1 vocalic release does not facilitate the recovery of C1

To test whether the vocalic release (transitional vocoid) helps the recovery of C1 gesture, we tested (1) whether the tokens with vocalic releases (*Presence of Vocalic Release*) were better identified, and (2) for the tokens with the vocalic releases, whether the releases of longer duration (*Duration of Vocalic Release*) led to better identification. The statistical models followed the same structure as in Section 6.1, except that C1 voicing was not considered, as none of the tokens with voiceless C1 included vocalic releases. Also, the vocalic release duration model was fitted to the subset of the data which contains only the stimuli produced with vocalic releases ( $n = 2,101$ , among which 1,148 was back-to-front. The entire dataset included 3,107 observations). **Table 5** shows the best models for the two measures.

Measure	Best mixed-effect logistic regression model
Presence of vocalic release	<b>Presence of Vocalic Release</b> * Place Order + (1 + Place Order + Presence of Vocalic Release  Subject) + (1  Item)
Vocalic release duration	<b>Vocalic Release Duration</b> + Place Order + (1 + Vocalic Release Duration + Place Order  Subject) + (1 Item)

Table 5: Best models for each measure.

First, the presence of a vocalic release influenced the C1 identification, but the effect interacted with Place Order [ $\chi^2(1) = 18.01, p < 0.0001^{***}$ ], as shown in **Figure 9**. A post-hoc pairwise comparison on the significant interaction revealed that, in back-to-front sequences only, C1 in the tokens without vocalic releases was better identified than C1 in those tokens with vocalic releases [ $\beta = 2.089, z = 3.97, p = 0.0001^{***}$ ]. For front-to-back sequences, presence versus absence of vocalic releases did not influence C1 identification [ $p = 0.55$ ]. Overall, the presence of a vocalic release did not improve C1 identification.



**Figure 9:** Predicted C1 ID accuracy by presence/absence of vocalic release (transitional vocoid).

For the tokens produced with the vocalic releases, the duration of these releases was a significant predictor of C1 identification regardless of the Place Order [ $\chi^2(1) = 10.73, p = 0.0011^{**}$ ]. However, the duration of vocalic releases and the C1 identification were inversely related as shown in **Figure 10**. That is, as the duration of the releases increased, the identification of C1 became worse [ $\beta = -0.056, z = -3.41, p = 0.0007^{***}$ ]. Participants inaccurately transcribed C1 when the releases were longer.

In sum, the outcomes of the two models related to the vocalic release clearly show that the vocalic releases do not facilitate the correct identification of C1. Rather, such transitional vocoids may interfere with the identification of the intended C1, especially when they are longer.

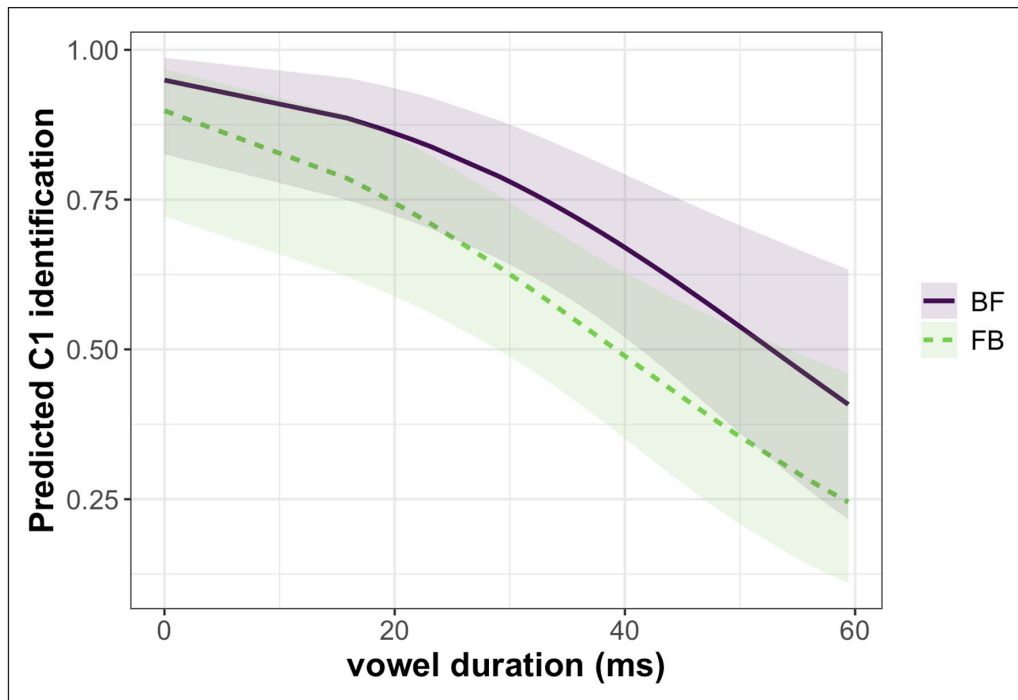


Figure 10: Predicted C1 ID accuracy by vocalic release duration (ms).

## 7. Discussion

Taken together, the results support the view that temporal organization, rather than vocalic releases, facilitates the perceptual recovery of C1. In the following sections, we discuss the implications of these findings, focusing first on the effects of temporal lag and overlap (Section 7.1), and then on the role of vocalic releases (Section 7.2).

### 7.1 Timing lag

The current findings support the first hypothesis that longer timing lag and reduced overlap between C1 and C2 facilitate the recovery of C1 gestures: Georgian native listeners do benefit from longer lag between C1 and C2 in recovering C1 in stop-stop sequences. Among the four measures tested, *Acoustic Lag*, *Onset Lag*, and *Relative Overlap*, each contributed to increasing accurate C1 identification, as predicted. Among these, *Acoustic Lag* and *Onset Lag* are moderately correlated.

The fourth lag measure we considered—*Release Lag*—does not correlate with *Acoustic Lag* and shows only a weak correlation with *Onset Lag*. Our results showed an interaction between *Release Lag* and place order, indicating that the influence of release lag differs significantly in B-F and F-B sequences. In B-F sequences, C1 accuracy did improve with longer release lags, but this was not the case for F-B sequences.

Overall, the results concerning timing lag are consistent with H1. Georgian native listeners identified C1 more accurately in sequences with reduced overlap, when the gesture for C2 is timed later relative to C1. This effect was evident in the timing of the C2 gesture onset for all the stop-stop sequences considered. The later timing of the C2 release proved beneficial for the B-F sequences (e.g., /gb/, /t<sup>h</sup>b/), presumably because the posterior-to-anterior transition increases the chance that a C1 release is masked. Release Lag had no significant impact on F-B sequences, suggesting that the perceptual vulnerability of C1 may vary by place order. These results thus align with the production patterns observed in previous studies (e.g., Chitoran et al., 2002), suggesting that native speakers may draw on perceptual knowledge when timing consecutive constrictions in onset clusters. The current findings, taken together with the previous findings on production patterns, provide converging evidence that timing lags (especially those based on gestural onsets) play a considerable role in recovering consonantal gestures. This suggests that the inter-consonantal timing lags within onset clusters are not merely a phonetic artefact but a part of grammar functionally encoded by Georgian speakers.

It is worth mentioning at this point that the overlap measures we considered crucially rely on gestural releases. This includes the acoustic measure we used—the inter-burst interval. Is it possible, though, that listeners may use other types of acoustic information that would correspond to consonant overlap? One possibility we have not considered, but has been proposed by Zsiga (2003), is duration ratio, a relative measure defined as the mean closure duration of the C1C2 cluster (with or without an intervening release) divided by the sum of the mean closure durations of single, intervocalic C1 and C2. This has been a useful measure of overlap across word boundaries in comparing non-native productions (English speakers producing L2 Russian utterances vs. Russian speakers producing L2 English utterances). It is worth verifying in the future whether this relative acoustic measure may be better correlated with the articulatory overlap measures.

For now, however, since C-C coordination with reduced overlap is known to often result in a vocalic release (transitional vocoid), especially in voiced sequences, we turn to the question of whether the presence of such vocalic releases benefits the recoverability of C1.

## 7.2 Vocalic releases

The second hypothesis, that the presence of C1 vocalic releases with richer C1 formant transition information would help listeners recover C1, is not supported by our results. Vocalic releases instead seem to be detrimental. This suggests that, unlike timing lag or degree of overlap, the presence of vocalic releases between two stops in C1C2 sequences may not be related to perceptual considerations. Or, if native speakers of Georgian indeed produce them for the listeners' sake, this result suggests that their efforts are ineffective—at least in the stop-stop sequences examined in this study.

An interaction with Place order was present in the results related to vocalic releases, as well. In F-B stop-stop sequences, the presence of a vocalic release had no effect on C1 identification. In B-F sequences, contrary to expectations, tokens without a vocalic release were more accurately identified than those with a vocalic release, a result which is not consistent with H2.

The duration of the vocalic release, when present, emerged as a significant predictor of accuracy, but again, contrary to H2, as the vocoid duration increases, C1 identification actually decreases. The inverse correlation between vocalic release duration and perceptual accuracy of C1 is at odds with findings from previous cross-language studies. For example, Wilson et al. (2014) and Wilson and Davidson (2013) found that longer burst duration resulted in more epenthesis between C1 and C2 stops when shadowing foreign speech, proposing that a longer burst has a similar acoustic profile to that of a vowel. In their study, greater burst amplitude similarly increased the rate of epenthesis. Both longer burst duration and greater burst amplitude are interpretable as presence of a vocoid, which protects C1 from misperception. However, these were the results of non-native listeners (English listeners hearing Russian stimuli), whose native phonotactics led them to interpret the acoustic properties of a longer and louder burst as a vowel. Georgian listeners confronted with non-native French data (Kwon & Chitoran, 2024) behaved differently because of their different native phonotactic expectation. Used to the presence of a vocalic release in a CC sequence, they did not reliably distinguish French CCV vs. CVCV stimuli in a non-native AX discrimination task. This result and the current findings together suggest that the quality and duration of V1 relative to V2 matters. When V1 is a vocalic release (transitional vocoid) and significantly shorter than V2, a lexical vowel, Georgian listeners tend to ignore it. However, a longer vocoid, one comparable in duration with the lexical vowel but not in its spectral properties, has a negative effect on C1 identification. Georgian listeners are thus sensitive to a longer vocoid, but it provides them with misleading information.

A possible reason for this particular result may be that the presence of a long vocalic release interferes with the perception and recoverability of information at multiple structural levels, not just the segmental level. It can simultaneously affect the perceptibility of C1, as well as the perceptibility of higher-level prosodic structures, such as syllables or words, thus impinging on overall intelligibility. A vocalic release may be perceived as a full vowel, therefore a syllable nucleus. In support of this interpretation, Crouch et al. (2023b) found that transitional vocoids in Georgian C1C2 sequences alter the amplitude envelope in ways that may lead to resyllabification. In sequences with sonority plateaus or falls, productions with a vocoid showed an additional peak in the amplitude envelope, compared to productions of the same sequence without a vocoid. This additional peak can be interpreted as a syllable nucleus and may lead listeners to mis-parse syllable boundaries, particularly when the vocoid is long enough to be perceptually salient but not clearly vowel-like. This suggests that speakers may avoid vocoids in certain cluster types not to preserve segmental clarity but to avoid prosodic ambiguity. If so, then

phonological grammar may encode not just segmental recoverability but also the recoverability of the syllabic organization. These findings are relevant for our current study, in which longer vocalic releases led to more CV responses and proved not effective for C1 identification.

So far, we (and other authors before us) have only considered segment-level perceptibility, with a focus on the recoverability of phonological contrasts. But we must consider the possibility that patterns that are predicted to facilitate the recovery of segmental contrasts may not be equally beneficial, and may even hinder, the recovery of syllable- or higher-level information.

### 7.3 Conclusion

We now return to our two initial hypotheses to consider what we have learned. Do the observed patterns—timing lag and vocalic releases—truly present an advantage for the recoverability of phonological information? Based on our results, we can maintain that the observed lag and overlap patterns in stop-stop sequences indeed help recover the phonological identity of C1, supporting H1. Vocalic releases, however, do not, contra H2. This is explained by the finding that the presence of the vocalic release is not linearly correlated with lag and thus does not provide a reliable cue to temporal organization. It may be used, instead, for conveying other types of information, such as voicing or cluster type.

Unlike prior studies that inferred timing-based perceptual effects from production data or nonnative perception data, the current findings offer direct perceptual evidence from native listeners. We provide evidence that timing lag matters in speech perception, contributing to segmental recovery in consonant clusters. These findings suggest that the native listeners rely on timing lag as the phonetic details relevant for recovering consonantal gestures. Yet the lack of perceptual benefit from vocalic releases suggests that not all phonetic patterns assumed to enhance recoverability serve that function. This contrast underscores the importance of evaluating perceptual recoverability empirically, rather than assuming their perceptual efficacy based on production patterns alone.

Corroborating previous findings on Georgian productions (Chitoran et al., 2002; Crouch, 2022), the current results support the inclusion of timing lag in the grammar of Georgian. Listeners benefit from longer lag (i.e., reduced overlap) and speakers tend to produce longer lag when C1 is expected to be more perceptually vulnerable, though it remains unclear whether speakers are manipulating the lag to aid listeners.

Going back to Mattingly (1981), we may ask: Do the speakers' phonological grammar integrate knowledge about temporal patterns and their acoustic consequences? The different timing lags in back-to-front and front-to-back sequences observed in production may be perceptually motivated in stop-stop sequences and learned as grammatical generalizations. Additional factors, not related to perceptibility, may also contribute to timing patterns. Pouplier et al.'s (2022) comparison of

lag patterns in the same CC clusters across seven languages firmly highlighted wide language-specific diversity, and, at the same time, consistent lag patterns across languages in terms of the segmental composition of the clusters. Among the seven languages, Georgian stands out as having the largest variability of lag durations. It is the language that reaches the longest lag durations, but in other respects it also conforms to cross-linguistic cluster-specific patterns. For example, the Georgian /sC/ clusters align with those of the six other languages in having the shortest lag. The large within-language variability of lag in Georgian is consistent with our conclusion that timing lag is part of speakers' knowledge. Controlling lag allows the recovery of many types of onset clusters, some of which, importantly, contain morphological information. Georgian is a language with rich prefixal morphology, where multiple consonantal prefixes may be added. This particular aspect of the language structure highlights the importance of both segmental and prosodic recoverability. A consonantal prefix must be accurately recovered and, at the same time, adding consonantal prefixes must not alter the syllable count. The precise control of inter-consonantal timing lag, avoiding long vocalic releases, is therefore important for these closely intertwined intelligibility considerations.

The current findings suggest that perceptual recoverability should not be defined strictly at a local, segmental level. A more fruitful approach should consider multiple levels of structures simultaneously—segments, syllabic organization, morphological structure—in the context of the overall intelligibility of the message. Speakers likely aim to be understood and to accommodate the listener, but this intent may be reflected in the phonological grammar in a complex way.

A further important point that needs to be raised is whether diachronic information should also be considered when examining timing patterns. Easterday (2019) raised this issue specifically regarding obstruent sequences in languages classified as having a highly complex syllable structure. She notes that the most common typological source of clusters is vowel reduction, and some characteristics of the reduced vowel may be retained in C1 release bursts. From a diachronic perspective then, the observed place effects may be seen as motivated by perceptual recoverability or a consequence of perceptual recoverability. The timing patterns and perceptual properties of such clusters may have the effect of preserving complex syllable structure by protecting it from complete overlap that can lead to consonant loss. Georgian stop-stop sequences are known to have two historical sources (Gamkrelidze & Ivanov, 1995). Back-to-front sequences developed through the deletion of an intervening vowel. In line with Easterday's reasoning, it can be argued that B-F sequences may have preserved the timing of a lost vowel. Front-to-back sequences with a dorsal C2 are known as 'harmonic clusters', and have developed from velarized stops, single segments that subsequently broke into sequences (e.g., [dʷ] becoming [dg] or [dɣ]). Along the same reasoning, they may have preserved a timing closer to that of a single segment. Such stability of timing patterns across time can be taken as further evidence for the phonological status of timing information.

While the current findings, together with previous studies on Georgian clusters, provide strong evidence for the perceptual role of timing lag in stop-stop sequences, further work is needed to determine whether similar mechanisms underlie the timing of other cluster types or hold cross-linguistically. Regardless of their perceptual motivation, the accumulating evidence in the literature indicates that timing differences constitute part of language-specific phonological knowledge. The role of timing in phonological grammar has indeed become increasingly clear across languages and different types of phonological contrasts. Gafos (2002) demonstrated the role of temporal coordination in Moroccan Arabic templatic word formation, becoming the first study to develop a full-fledged formal analysis incorporating gestural timing. Since then, the phonological role of *timing* (whether gestural or not) has found additional support, in particular in the instantiation of phonological contrasts. Tone contrasts in several languages, for example, have been shown to consist of tonal units that differ exclusively in their timing (e.g., Remijsen & Ayoker, 2014, for Shilluk; Svensson Lundmark et al., 2021, for Swedish; Karlin, 2022, for Serbian). Segmental contrast between complex segments and sequences of segments has been shown to rely on the different timing of the same component articulatory gestures (Shaw et al., 2021).

Our view of the role of timing in phonology, as supported by the Georgian data examined here, most closely resembles the one presented by Gafos et al. (2020), in three aspects:

1. We argue, on the basis of our results, that inter-gestural timing patterns and their perceptual relevance can show language-specificity. The Georgian timing patterns on which native listeners rely perceptually are specific to Georgian in the same way that inter-segmental temporal coordination is shown to be language-specific in Gafos et al. (2020). In the latter case, differences in temporal coordination account for language-specific differences in syllable affiliation between Arabic and Spanish. In the case of Georgian clusters, Kwon and Chitoran (2024) show that French listeners' perception differs from Georgian listeners, suggesting further evidence for the language-specificity in the domain of perception.
2. It is the timing pattern, and not the presence or absence of a vocalic release, that is part of native speakers' phonological knowledge.
3. How speakers organize their vocal tracts is not independent of how they organize their native linguistic system in their minds (Gafos et al., 2020). Speakers of Georgian and Arabic have to accommodate morphological structure that impinges on prosodic and segmental requirements. When these requirements conflict is when we are likely to see exceptions from typological generalizations (e.g., the blatant disregard for segmental combinations that follow the sonority sequencing principle, in both Georgian and Arabic).

The results of our own study provide perceptually motivated explanations for the wide variability of overlap patterns in Georgian in Pouplier et al. (2022). The consonant sequences compared in

that study were exclusively those that occurred in all of the seven languages under investigation (obstruent-liquid, sibilant-obstruent, and /kn/, /gn/, for a subset of the languages). A plausible interpretation for the reduced overlap measures found in some of the Georgian data is that they are motivated by the presence of stop-stop sequences in the language, the sequences that require close control of timing for recoverability reasons. The variable timing lag in Georgian, expanding further into the reduced overlap range than the other languages, can be seen as an optimization solution for the temporal unfolding of the component gestures. Importantly, reduced overlap (long lag) is not simply generalized across all CC sequences in Georgian. If it were, then presumably a sequence like /sp/ would no longer be perceived as two adjacent consonants. Instead, a broader overlap range is the preferable compromise such that, depending on their gestural composition, sequences spread out between the increased overlap (shorter lag) range, and reduced overlap (longer lag).

To sum up, the current study provides perceptual evidence for incorporating timing lag into phonological representations. The perceptual evidence has explanatory value, as it allows us to probe the relationship between vocal tract organization and linguistic knowledge. Taken together with the articulatory data (Pouplier et al., 2022), the current findings offer us a glimpse of how speakers manipulate vocal tract organization reflecting linguistic structure, and how listeners, in turn, use the temporal properties to recover the hierarchical, as well as the segmental, information.

---

## Additional file

The additional file for this article can be found as follows:

- **Supplementary Materials.** A. Acoustic and Articulatory Measurements of Perception Stimuli and B. Statistical Model Output. <https://doi.org/10.16995/labphon.24395.s1>

## Acknowledgements

We thank the Georgian speakers and experiment participants in Tbilisi, the anonymous reviewers for their valuable comments, and Louis Goldstein and Marianne Pouplier for helpful discussions. A previous version of this study was presented at LabPhon 16, and benefitted from insightful comments by Jennifer Hay, the discussant.

## Funding information

This work was supported by ANR-DFG grant (ANR-14-FRAL-0004) for the project PATHS, and by the IdEx programme (ANR-18-IDEX-0001) to Université Paris Cité. Research in Georgia was made possible by the Fulbright-Hays programme of the US Department of State.

## Competing interests

The authors have no competing interests to declare.

## Author contributions

IC: Conceptualization, study design, funding acquisition, data collection, interpretation, writing, editing.

HK: Data analysis, interpretation, writing, editing.

---

## References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4), 785–821. <https://doi.org/10.1353/lan.0.0165>
- Benki, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica*, 60(2), 129–157. <https://doi.org/10.1159/000071450>
- Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [Computer program]. Version 6.1.50. <http://www.praat.org/>
- Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston, & M. E. Beckman (Eds.), *Papers in Laboratory Phonology* (1<sup>st</sup> ed., pp. 341–376). Cambridge University Press. <https://doi.org/10.1017/CBO9780511627736.019>

- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180. <https://doi.org/10.1159/000261913>
- Byrd, D. (1992). Perception of assimilation in consonant clusters: A gestural model. *Phonetica*, 49, 1–24. <https://doi.org/10.1159/000261900>
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209–244. <https://doi.org/10.1006/jpho.1996.0012>
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modelling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149–180. [https://doi.org/10.1016/S0095-4470\(02\)00085-2](https://doi.org/10.1016/S0095-4470(02)00085-2)
- Chen, L. H. (2003). Evidence for the role of gestural overlap in consonant place assimilation. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences*, Barcelona, Spain, August 3–9, 2003. <http://www.internationalphoneticassociation.org/icphs/icphs2003>
- Chitoran, I., & Goldstein, L. (2006). *Testing the phonological status of perceptual recoverability: Articulatory evidence from Georgian*. Abstract, LabPhon 10, June 29–July 1, Paris, France.
- Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology*, 7, pp. 419–447. <https://doi.org/10.1515/9783110197105.2.419>
- Cho, T., Yoon, Y., & Kim, S. (2014). Effects of prosodic boundary and syllable structure on the temporal realization of CV gestures. *Journal of Phonetics*, 44, 96–100. <https://doi.org/10.1016/j.wocn.2014.02.007>
- Crouch, C. (2022). *Postcards from the syllable edge: Sonority and articulatory timing in complex onsets in Georgian*. [Doctoral dissertation, University of California, Santa Barbara]. <https://escholarship.org/uc/item/5w18167d>
- Crouch, C., Chitoran, I., Goldstein, L., & Katsika, A. (2023b). Intrusive vocoids and syllable structure in Georgian. In R. Skarnitzl, & J. Volín (Eds.), *Proceedings of the 20<sup>th</sup> International Congress of Phonetic Sciences – ICPhS* (pp. 2000–2004). August 7–11, 2023. Prague, Czech Republic. [https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full\\_papers/613.pdf](https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/613.pdf)
- Crouch, C., Katsika, A., Chitoran, I. (2023a). Sonority sequencing and its relationship to articulatory timing in Georgian. *Journal of the International Phonetic Association*, 53(3). <https://doi.org/10.1017/S0025100323000026>
- Dell, F., & Elmedlaoui, M. (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-010-0279-0>
- Du, S., & Gafos, A. (2022). Articulatory overlap as a function of stiffness in German, English and Spanish word-initial stop-lateral clusters. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 14(1). <https://doi.org/10.16995/labphon.7965>
- Easterday, S. (2019). Highly complex syllable structure. A typological and diachronic study. *Studies in Laboratory Phonology*, 9. Language Science Press.
- Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89(1). <https://doi.org/10.1121/1.400674>

- Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, 20, 269–337. <https://doi.org/10.1023/A:1014942312445>
- Gafos, A. I., Hoole, P., Roon, K. D., & Zeroual, C. (2010). Variation in timing and phonological grammar in Moroccan Arabic clusters. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology*, 10 (pp. 657–698). Mouton de Gruyter. <https://doi.org/10.1515/9783110224917.5.657>
- Gafos, A. I., Roeser, J., Sotiropoulou, S., Hoole, P., & Zeroual, C. (2020). Structure in mind, structure in vocal tract. *Natural Language & Linguistic Theory*, 38(1), 43–75. <https://doi.org/10.1007/s11049-019-09445-y>
- 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). Mouton de Gruyter. <https://doi.org/10.1515/9783110815030>
- Goldstein, L., & Fowler, C. (2003). Articulatory Phonology: A phonology for public language use. In N. Schiller, & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production* (pp. 159–207). Mouton de Gruyter. <https://doi.org/10.1515/9783110895094.159>
- Hall, N. (2024). Intrusive and epenthetic vowels revisited. In J. Y. Kim, V. Miatto, A. Petrović, & L. Repetti (Eds.), *Epenthesis and beyond: Recent approaches to insertion in phonology and its interfaces* (pp. 167–197). Language Science Press.
- Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, 4, 247–263. [https://doi.org/10.1016/0167-6393\(85\)90051-2](https://doi.org/10.1016/0167-6393(85)90051-2)
- Hardcastle, W. J., & Roach, P. (1979). An instrumental investigation of coarticulation in stop consonant sequences. In H. Hollien, & P. Hollien (Eds.), *Current issues in the phonetic sciences* (pp. 531–540). John Benjamins. <https://doi.org/10.1075/cilt.9.56har>
- Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, 123, 2825–2835. <https://doi.org/10.1121/1.2897042>
- Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2019). The phonetic basis of the origin and spread of sound change. In W. F. Katz, & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 401–426). Routledge. <https://doi.org/10.4324/9780429056253-15>
- Harris, A. (2002). The word in Georgian. In Dixon, R., & A. Aikhenvald (Eds.), *Word: A cross-linguistic typology* (pp. 127–142). Cambridge University Press.
- Hayes, B., Kirchner, R., & Steriade, D. (Eds.) (2004). *Phonetically based phonology*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486401>
- Henke, E., Kaisse, E. M., & Wright, R. (2012). Is the Sonority Sequencing Principle an epiphenomenon? In Parker, S. (Ed.), *The sonority controversy* (pp. 65–100). Mouton De Gruyter. <https://doi.org/10.1515/9783110261523.65>
- Iskarous, K. (2017). The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics*, 64, 8–20. <https://doi.org/10.1016/j.wocn.2017.05.003>

- Iskarous, K., & Kavitskaya, D. (2018). Sound change and the structure of synchronic variability: Phonetic and phonological factors in Slavic palatalization. *Language*, 94(1), 43–83. <https://doi.org/10.1353/lan.2018.0001>
- Karlin, R. (2022). Expanding the gestural model of lexical tone: Evidence from two dialects of Serbian. *Journal of Laboratory Phonology*, 13(1). <https://doi.org/10.16995/labphon.6443>
- Katsika, A. (2016). The role of prominence in determining the scope of boundary lengthening in Greek. *Journal of Phonetics*, 55, 149–181. <https://doi.org/10.1016/j.wocn.2015.12.003>
- Katz, W. F., Bharadwaj, S. V., & Stettler, M. P. (2006). Influences of electromagnetic articulography sensors on speech produced by healthy adults and individuals with aphasia and apraxia. *Journal of Speech, Language, and Hearing Research*, 49(3), 645–659. [https://doi.org/10.1044/1092-4388\(2006/047\)](https://doi.org/10.1044/1092-4388(2006/047))
- Kochetov, A., & Goldstein, L. (2005). Position and place effect in Russian word-initial and word-medial stop clusters. *Journal of the Acoustic Society of America*, 117(4), 2571. <https://doi.org/10.1121/1.4788568>
- Kochetov, A., Pouplier, M., & Son, M. (2007). Cross-language differences in overlap and assimilation patterns in Korean and Russian. *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences* (pp. 1361–1364), Saarbrücken.
- Kochetov, A., & So, C. K. (2007). Place assimilation and phonetic grounding: A cross-linguistic perceptual study. *Phonology*, 24(3), 397–432. <https://doi.org/10.1017/S0952675707001273>
- Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, 35, 162–179. <https://doi.org/10.1016/j.wocn.2006.04.001>
- Kühnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. *Proceedings of the 7<sup>th</sup> International Seminar on Speech Production*. (pp. 327–334).
- Kwon, H., & Chitoran, I. (2024). Perception of illusory clusters: The role of native timing. *Phonetica*. <https://doi.org/10.1515/phon-2023-2005>
- Lenth, R. V. (2022). emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.3, <<https://CRAN.R-project.org/package=emmeans>>
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 576–585. <https://doi.org/10.1037/0096-1523.15.3.576>
- Mattingly, 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. [https://doi.org/10.1016/S0166-4115\(08\)60217-4](https://doi.org/10.1016/S0166-4115(08)60217-4)
- Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from a Parasession on Language and Behavior* (pp. 178–203). Chicago Linguistics Society.
- Peng, S.-H. (1996). *Phonetic implementation and perception of place coarticulation and tone sandhi*. [Doctoral dissertation, Ohio State University]. [http://rave.ohiolink.edu/etdc/view?acc\\_num=osu1384525774](http://rave.ohiolink.edu/etdc/view?acc_num=osu1384525774)

- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92(6), 3078–3096. <https://doi.org/10.1121/1.404204>
- Poupplier, M., Marin, S., Hoole, P., & Kochetov, A. (2017). Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics*, 64, 108–126. <https://doi.org/10.1016/j.wocn.2017.01.006>
- Poupplier, M., Pastätter, M., Hoole, P., Marin, S., Chitoran, I., Lentz, T. O., & Kochetov, A. (2022). Language and cluster-specific effects in the timing of onset consonant sequences in seven languages. *Journal of Phonetics*, 93. <https://doi.org/10.1016/j.wocn.2022.101153>
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Remijsen, B., & Ayoker, O. G. (2014). Contrastive tonal alignment in falling contours in Shilluk. *Phonology*, 31(3), 435–462. <https://doi.org/10.1017/S0952675714000219>
- Ridouane, R. (2016). Leading issues in Tashlhiyt phonology. *Language and Linguistics Compass*, 10(11), 644–660. <https://doi.org/10.1111/lnc3.12211>
- Ridouane, R., & Cooper-Leavitt, J. (2019). A story of two schwas: A production study from Tashlhiyt. *Phonology*, 36(3), 433–456. <https://doi.org/10.1017/S0952675719000216>
- Ridouane, R., & Fougeron, C. (2011). Schwa elements in Tashlhiyt word-initial clusters. *Journal of Laboratory Phonology*, 2, 275–300. <https://doi.org/10.1515/labphon.2011.010>
- Ridouane, R., Hermes, A., & Hallé, P. (2014). Tashlhiyt's ban of complex syllable onsets: Phonetic and perceptual evidence. *STUF – Language Typology and Universals*, 67(1), 7–20. <https://doi.org/10.1515/stuf-2014-0002>
- Roon, K. D., Hoole, P., Zeroual, C., Du, S., & Gafos, A. I. (2021). Stiffness and articulatory overlap in Moroccan Arabic consonant clusters. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 12(1), 8. <https://doi.org/10.5334/labphon.272>
- Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, 70, 321–328. <https://doi.org/10.1121/1.386780>
- Shaw, J., Oh, S., Durvasula, K., & Kochetov, A. (2021). Articulatory coordination distinguishes complex segments from segment sequences. *Phonology*, 38(3), 437–477. <https://doi.org/10.1017/S0952675721000269>
- Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>
- Son, M. (2008). Gradient reduction of C1 in /pk/ sequences. *Phonetic Sciences*, 15(4), 43–65.
- Sorensen, T., & Gafos, A. (2016). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology*, 28(4), 188–215. <https://doi.org/10.1080/10407413.2016.1230368>
- Sotiropoulou, S., & Gafos, A. (2022). Phonetic indices of syllabic organization in German stop-lateral clusters. *Journal of the Association for Laboratory Phonology*, 13(1), 1–42. <https://doi.org/10.16995/labphon.6440>

- Steriade, D. (2008). The phonology of perceptibility effects: The P-Map and its consequences for constraint organization. In K. Hanson, & S. Inkelas (Eds.), *The nature of the word: Studies in honor of Paul Kiparsky* (pp. 150–179). MIT Press. <https://doi.org/10.7551/mitpress/9780262083799.003.0007>
- Surprenant, A., & Goldstein, L. (1998). The perception of speech gestures. *Journal of the Acoustical Society of America*, 104(1), 518–529. <https://doi.org/10.1121/1.423253>
- Svensson Lundmark, M., Frid, J., Ambrazaitis, G., & Schötz, S. (2021). Word-initial consonant-vowel coordination in a lexical pitch-accent language. *Phonetica*, 78(5–6), 515–569. <https://doi.org/10.1515/phon-2021-2014>
- Tienkamp, T. B., Rebernik, T., Jacobi, J., Wieling, M., & Abur, D. (2024). The impact of electromagnetic articulography sensors on the articulatory-acoustic vowel space in speakers with and without Parkinson's disease. *Proceedings of the 13<sup>th</sup> International Seminar on Speech Production*. 13–17 May 2024, Autrans, France. <https://doi.org/10.21437/issp.2024-24>
- Wilson, C., & Davidson, L. (2013). Bayesian analysis of non-native cluster production. In S. Kan, C. Moore-Cantwell, & R. Staubs (Eds.), *Proceedings of NELS 40*. Amherst, MA: Graduate Linguistics Student Association (pp. 265–278).
- Wilson, C., Davidson, L., & Martin, S. (2014). Effects of acoustic-phonetic detail on cross-language speech production. *Journal of Memory and Language*, 77, 1–24. <https://doi.org/10.1016/j.jml.2014.08.001>
- Wright, R. (1996). *Consonant clusters and cue preservation in Tsou*. [Doctoral dissertation, University of California, Los Angeles]. <https://linguistics.ucla.edu/images/stories/wright.1996.pdf>
- Wright, R. (2001). Perceptual cues in contrast maintenance. In K. Johnson, & E. Hume (Eds.), *The role of speech perception in phonology* (pp. 251–277). Brill. [https://doi.org/10.1163/9789004454095\\_014](https://doi.org/10.1163/9789004454095_014)
- Yanagawa, M. (2006). *Articulatory timing in first and second language: A cross-linguistic study*. [Doctoral dissertation, Yale University].
- Yip, J. C. K. (2013). *Phonetic effects on the timing of gestural coordination in Modern Greek consonant clusters*. [Doctoral dissertation, University of Michigan]. <https://www.proquest.com/docview/1497967202>
- Zellou, G., Lahrouchi, M., & Bensoukas, K. (2024). The perception of vowelless words in Tashlhiyt. *Glossa: A journal of general linguistics*, 8(1), 1–41. <https://doi.org/10.16995/glossa.10438>
- Zhgenti, S. (1956). *Kartuli enis 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. [https://doi.org/10.1016/S0095-4470\(19\)30189-5](https://doi.org/10.1016/S0095-4470(19)30189-5)
- Zsiga, E. C. (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, 28, 69–102. <https://doi.org/10.1006/jpho.2000.0109>
- Zsiga, E. C. (2003). Articulatory timing in a second language: Evidence from Russian and English. *Studies in Second Language Acquisition*, 25, 399–432. <https://doi.org/10.1017/S0272263103000160>

