



Perception of illusory clusters: the role of native timing

Harim Kwon, Ioana Chitoran

► To cite this version:

Harim Kwon, Ioana Chitoran. Perception of illusory clusters: the role of native timing. *Phonetica*, 2023, 10.1515/phon-2023-2005 . hal-04312761

HAL Id: hal-04312761

<https://hal.science/hal-04312761>

Submitted on 28 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Research Article

Harim Kwon* and Ioana Chitoran

Perception of illusory clusters: the role of native timing

<https://doi.org/10.1515/phon-2023-2005>

Published online November 29, 2023

Abstract: We explore the influence of native timing patterns on nonnative speech perception, by asking whether a nonnative CVCV sequence can be perceived as CCV when the temporal organization of nonnative CVCV is similar to native CCV. To explore this question, Georgian listeners are tested on a CCa-CVCá discrimination in French. Georgian has a rich word-onset cluster inventory, with component consonants loosely timed. The loose timing often, though not always, results in a schwa-like CC transition. French, the stimulus language, exhibits tighter timing in biconsonantal clusters, no vocalic transitions, and a reduced non-prominent first vowel in CVCá sequences. We hypothesize that the cross-language difference in inter-consonantal timing can facilitate the perception of an illusory cluster when Georgian listeners hear French CVCá. The findings reveal such perceptual confusion, particularly in the CCa-CøCá contrast in which the nonnative /ø/ is phonetically similar to the CC transition in Georgian, both in terms of temporal organizations and tongue shape. This confirms the possibility of illusory clusters, which is consistent with the interpretation that Georgian listeners utilize their knowledge of how word-onset CC clusters are temporally implemented in their native language when responding to the task. We propose that the timing pattern may constitute language-specific knowledge and that it can influence the perceptual assimilation patterns in nonnative speech perception.

Keywords: nonnative perception; illusory cluster; temporal organization; Perceptual Assimilation Model

***Corresponding author: Harim Kwon**, Department of English Language and Literature, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea, E-mail: harimkwon@snu.ac.kr

Ioana Chitoran, Department of Linguistics, Clillac-ARP, Université Paris Cité, Paris, France, E-mail: ioana.chitoran@u-paris.fr

1 Introduction

1.1 Background

Languages differ in how they time successive consonants in a sequence (e.g., Bombien and Hoole 2013; Hoole et al. 2009; Kochetov et al. 2007; Pouplier et al. 2020, 2022; Zsiga 2003). In the case of word-onset consonant clusters, for instance, the component consonants of biconsonantal onset clusters are more tightly timed in German than those in Georgian (Pouplier et al. 2020) or French (Bombien and Hoole 2013). These differences in timing pose difficulties in second language learning (Zsiga 2003) or imitating nonnative sequences (Pouplier et al. 2020), and have been claimed to be part of native speakers' knowledge that is representational and language-specific (e.g., Gafos 2002; Gafos and Goldstein 2012).

Georgian is a language in which the component consonants are loosely timed in word-onset clusters (e.g., Chitoran et al. 2002; Crouch 2022; Pouplier et al. 2020). The loose timing between the component consonants, or the long inter-consonantal lag, often results in a transitional vocoid which is typically transcribed as a schwa. Both the temporal distance between the component consonants and the existence of the vocoid seem to be contextually conditioned. The inter-consonantal timing is conditioned by the composition of the consonant cluster. For example, in a bi-consonantal C_1C_2 cluster, the temporal distance between C_1 and C_2 is longer when the place of articulation is further back in C_1 than in C_2 (back-to-front order) than vice versa (front-to-back order) (e.g., Chitoran et al. 2002). Also, a sibilant C_1 tends to be more tightly timed with C_2 than a stop C_1 (Pouplier et al. 2022). The occurrence of the transitional vocoid also varies, and the vocoid is more likely to appear when the inter-consonantal lag is long and when the ambient consonants are voiced (Crouch 2022; Crouch et al. 2023a).

The extent of the timing lag and the presence of the vocoid are not independent (in fact, the latter is presumably a by-product of the former, Crouch et al. 2023b), but their variation seems to have different motivations. First, the variation in timing is systematic and, at least partially, attributable to perceptual considerations related to the composition of the clusters. When the component consonants are tightly timed with a short lag, C_1 in back-to-front clusters (e.g., /gb/) become perceptually less recoverable than in front-to-back clusters (e.g., /bg/), because, in a back-to-front cluster, a fronter closure (e.g., /b/) that overlaps with a preceding backer closure (e.g., /g/) can mask the release of the first consonant (Chitoran et al. 2002). If speakers are (subconsciously) aware of this consequence of tight timing, they may organize the consonantal gestures accordingly in back-to-front and front-to-back clusters. Tighter timing is tolerated when the manners of the component consonants make them less vulnerable to such

masking, as in the case of sibilant-initial clusters (Pouprier et al. 2022). On the contrary, when the manner of articulation could increase the adverse effects of tighter timing on perceptual recoverability, loose timing is preferred. Hoole et al. (2009) report tighter timing in /kl/ than in /kn/ in German, as a tightly timed nasal C₂ can jeopardize the release of /k/. This suggests that the temporal organization of the consonants may be controlled (subconsciously) by the speakers and it comprises a nontrivial part of their phonetic knowledge.

On the other hand, the variation in the presence or absence of the transitional vocoid seems to be more mechanical, as it is contingent, first, on the long inter-consonantal lag, and second, on the voicing of the consonants. The long lag between the component consonants would result in an open vocal tract, and when the vocal fold vibration is sustained throughout the successive consonants, it would give rise to a schwa-like vocoid (e.g., Davidson 2005). In Georgian, the appearance of the vocoids is correlated with the long inter-consonantal lag, but the duration of the lag (articulatorily measured) and that of the vocoid (acoustically measured) are not correlated (Crouch et al. 2023b). These suggest that the transitional vocoid within a consonant cluster is not an accurate acoustic correlate of the temporal organization of the component consonants though they are certainly related to each other.

A separate line of research has shown that speakers of a language that has a poor inventory of consonant clusters perceive a vowel in a sequence of consonants even when the signal does not include a vocalic element between the consonants (e.g., Berent et al. 2007, 2009; Davidson 2011; Dupoux et al. 1999). For example, Dupoux et al. (1999) show that Japanese listeners perceptually assimilate CC sequences, phonotactically illicit in Japanese, to CVC sequences. This illusory vowel, perceived without acoustic correlates in the signal, has been attributed to a perceptual “repair” of the sound sequences that are not allowed in the listeners’ native language. We will use the term “repair” to refer to “perceptual repair” that does not (necessarily) involve conscious computations based on the listeners’ phonological grammar, as used in the literature in loanword adaptation (for a review of the latter, see Kang 2011).

But the illusory vowel perception does not seem to be the only reason to cause the CCV-CVCV confusion. Though most previous studies have examined how listeners repair illicit CC sequences by inserting a vowel between the consonants to break them up (i.e., CCV is repaired into CVCV), it is logically possible that the vowel between the two consonants is perceptually deleted from CVCV (i.e., CVCV is repaired into CCV) if, somehow, CCV has a closer match in the listeners’ native language than CVCV does. This possibility has been suggested in Berent et al. (2009). Russian listeners in Berent et al. (2009) report hearing CCVC about 20 % of the time when the stimuli are English nasal-initial [CəCVC] (Berent et al. 2009, pp. 94–96). The authors attribute this to the rarity of pretonic schwas in Russian. The listeners perceptually modify an unfamiliar (or ungrammatical) CəCVC structure to a more familiar native

CCV structure. In addition to the culprit that Berent et al. (2009) identified (i.e., the rarity of pretonic schwas in Russian), we suggest other factors that may have contributed to this CVCV-to-CCV repair. In Russian, nasal-initial word onset clusters are well-formed and it is likely that the clusters involving nasals are among those produced with longer lags between consonants. We know from Pouplier et al. (2022) that this is cross-linguistically true at least for CN onset clusters. If long lag also characterizes the more rare NC onset clusters, this would make Russian NCV a good assimilatory target for English NəCV̌. We will refer to this kind of vowel deletion in nonnative perception as illusory cluster perception.

We hypothesize that illusory cluster perception can be facilitated by the loose timing or longer lag between the consonants composing a cluster in the listeners' native language. This study sets out to test this hypothesis with Georgian listeners. Georgian is a language with a rich word-onset cluster inventory, in which the component consonants of an onset biconsonantal cluster are loosely timed (e.g., Chitoran et al. 2002; Pouplier et al. 2020). French is used as the stimulus language as it provides an appropriate platform to test the illusory cluster perception in several different ways. First, French CVCV sequences have final prominence (e.g., Jun and Fougeron 2000; Vaissière 1983). Second, the non-prominent first vowels are reduced in their duration and quality (e.g., Adda-Decker et al. 2008; Meunier and Espesser 2011). Third, the French vowel /ə/, when reduced in a non-prominent syllable, is similar to schwa (e.g., Hall and Hume 2013). Lastly, consonants in an onset cluster are tightly timed in French (e.g., Pouplier et al. 2022), and thus, French CCV sequences do not typically involve a transitional vocoid. More details are in Section 1.3.

In the following sub-sections, we review how timing has been studied in conjunction with nonnative perception (Section 1.2), provide relevant background on Georgian and French (Section 1.3.), and present our research question and predictions (Section 1.4).

1.2 Timing in nonnative speech perception

Theoretical models of nonnative speech perception commonly predict that perception of nonnative sounds is influenced by the sound systems of the listeners' native language(s). According to PAM (Perceptual Assimilation Model, e.g., Best 1995), listeners are perceptually attuned to the phonetic (articulatory) differences that are linguistically significant (i.e., contributing to lexical/phonological contrasts) in their native language. The listeners' perception is streamlined as they become highly sensitive to contrastive phonetic properties and, at the same time, gradually lose sensitivity to the phonetic differences that are irrelevant to phonological contrasts. Native Language Magnet theory (e.g., Kuhl 1993), on the other hand, claims that

native language experience warps the perceptual space such that the listeners form the prototype of a certain sound category from the distributional properties of the language input. The prototypes function as magnets attracting the nearby sounds and the sounds that are attracted to the same magnet become less discriminable. Another theory, Automatic Selective Perception (Strange 2011), claims that listeners switch their perceptual routines according to the task. Instead of losing the ability to discriminate fine-grained phonetic details that are linguistically irrelevant, the listeners are selectively attending to the contrastive properties when the task is more complicated or realistic. On the contrary, when the task and the stimuli are simpler, listeners are more likely to turn their attention to small phonetic details that may not necessarily be linguistically relevant, increasing the possibility of detecting the details.

Of these different theories of nonnative speech perception, PAM explicitly indicates that perception works on the basis of articulatory gestures. Gestures are defined in terms of their temporal and spatial properties (e.g., Browman and Goldstein 1992), which makes PAM directly applicable to conceptualization of timing. In PAM, or in any other theories of speech production, perception, or phonology regarding gestures as the primitives, such as Direct Realism (Fowler 1996) or Articulatory Phonology (Browman and Goldstein 1992), segments are constellations of frequently co-occurring gestures that have spatial and temporal dimensions. Which gestures would constellate together to form a segment, as well as how the gestures are temporally organized, is language-specific and varies across languages. And this cross-language difference in gesture grouping and their timing would influence the perception of nonnative speech.

According to PAM, nonnative phones are assimilated to the sounds in the listeners' native language on the basis of the articulatory similarity, and the assimilation patterns predict the discriminability of a nonnative contrast. Specifically, the discrimination is near-ceiling if the contrast is perceived as equivalent to a native phonological contrast (Two Category assimilation), less high, but still good if the two members are perceived as a better and a poorer exemplar of the same category (Category Goodness difference), and lowest if both members are perceived as phonetically equivalent to a single native category (Single Category assimilation). For example, in Tyler et al.'s (2014) investigation of English monolingual listeners' perception of nonnative vowels, the Norwegian vowels /i/ and /y/ are both assimilated to /i/ whereas /ʉ/ is assimilated to /u/. The listeners' discrimination reflects this assimilation pattern such that /i/ and /ʉ/, but not /i/ and /y/, are accurately discriminated. The assimilation of the nonnative vowels seems to be determined by the articulatory similarities, in terms of the tongue configuration, between the vowel gestures in the signal and in the listeners' native language. For instance, English /i/

would be a close match to Norwegian /y/ in terms of the spatial similarities of the vowel gestures (i.e., where the highest position of the tongue is).

Best and Hallé (2010) show how the temporal organization of the gestures, in addition to their spatial properties, can influence the perceptual assimilation patterns. They investigate French and English listeners' perception of three temporally and spatially distinctive types of nonnative onset structures, Hebrew onset clusters /dl tɫ/, Zulu lateral fricatives /ɬ ɮ/, and Tlingit lateral affricates /dɬ ɬ/. Among these sounds, Zulu fricatives are often assimilated by listeners to complex onsets (e.g., stop + fricative sequences). This segment-to-cluster assimilation, according to Best and Hallé (2010), shows that the listeners perceptually repair the time structure of the unfamiliar sound in a way that is consistent with their native patterns. Zulu /ɬ/ could be assimilated, for instance, to a non-lateral (post-)alveolar fricative or affricate (repairs in the location and/or the degree of the constriction), but instead, it is more often assimilated to a two-segment sequence, regrouping the individual gestures involved in /ɬ/ into two separate segments. Based on these findings, Best and Hallé (2010) claim that the perceptual assimilation of nonnative speech needs to accommodate the time structure of the nonnative speech, especially when the temporal structure is the main difference between the nonnative speech sounds and the listeners' native language.

Gestures can be re-parsed (or 're-constellated') according to the commonly co-occurring patterns in the listeners' native language. Best and Hallé (2010) provide evidence for this gestural re-constellation due to temporal repair within syllabic onsets, which is a linguistically relevant unit. And they claim that the entire onset which may have one or more segments can be holistically perceived, and the articulatory gestures can be reorganized within the onset structure. In this study, we investigate the role of the temporal organization beyond the onset structure and test if a nonnative CVCV sequence can be repaired to CCV. We present a view that the CVCV-to-CCV assimilation, or the perception of illusory clusters, is due to the re-constellation of involved gestures into segments, in a similar way to the assimilation of nonnative affricates to stop + fricative sequences shown in Best and Hallé (2010). To this end, we test Georgian listeners' discrimination of French CCV sequences from CøCV which contains the nonnative phone /ø/ in its first syllable. According to PAM,¹ poor discrimination would mean that both sequences are perceptually assimilated to the same sequence in the listeners' native language. Greater discrimination accuracy

¹ In our view, PAM provides directly testable predictions and straightforward explanations with regard to the role of temporal information in nonnative speech perception, but we do not deny alternative explanations, nor do we aim to argue that the perception of illusory clusters can exclusively be explained by the phonetic theories based on articulatory gestures as the primitives of speech perception (see Section 1.4).

would suggest that the sequences are assimilated to different native sequences. If French CCV and CVCV are assimilated to the same Georgian sequence, it would either be CCV-to-CVCV assimilation or CVCV-to-CCV assimilation. We argue it is the latter because French CVCV and Georgian CCV can be quite similar in their temporal structure, as reviewed in the following section.

1.3 Test languages: Georgian and French

1.3.1 Word-onset consonant clusters

Georgian allows long consonant sequences, specifically in word-initial position (Aronson 1982, 1991; Butskhrikidze 2002; Hewitt 1995; Tschenkeli 1958; Vogt 1958). It is generally agreed in synchronic descriptions of Georgian, and confirmed by naïve native speaker intuitions, that words like [tkma] ‘to say’ and [ptskvna] ‘to peel’ are uncontroversially monosyllabic. Independent evidence can be found in Crouch (2022).

For the two-member clusters, Georgian allows more varied consonant combinations than French. The C_1C_2 combinations permitted in French comprise a proper subset of those allowed in Georgian. In addition to the phonotactic difference, Georgian and French also differ in how they implement the consonant clusters that are equivalent in their phonemic content. The most conspicuous differences between the two languages include the timing between two consonants within an onset CC cluster. In general, the component consonants in Georgian onset CC clusters are more loosely timed with longer inter-consonantal lag than those in French (e.g., Bombien and Hoole 2013; Kühnert et al. 2006; Pouplier et al. 2022). In addition to this cross-linguistic difference, Pouplier et al. (2022) demonstrate that the timing patterns are decided by the composition of the consonant clusters within a language. That is, individual languages have a general tendency toward tight or loose timing, but the composition of the cluster can interfere with this tendency. For instance, even in Georgian, sibilant-initial clusters exhibit as tight timing as in French, though the timing is much looser and more variable in other types of clusters (e.g., stop-initial clusters). This suggests that Georgian speakers would produce, and Georgian listeners would expect, a greater variation in the inter-consonantal timing patterns of word onset CC clusters, in comparison with French speakers who would be familiar only with the tight inter-consonantal timing.

This cross-linguistic timing difference between Georgian and French gives rise to another interesting variation, namely, a transitional vocoid. In Georgian onset CC clusters, a transitional vocoid is frequently observed between the two component consonants. The transitional vocoid is characterized as schwa-like and is related to

the relatively long timing lag between the consonantal gestures (e.g., Chitoran et al. 2002; Crouch et al. 2023b; Pouplier et al. 2020). Crouch et al. (2023b) have shown that, articulatorily, the transitional vocoids appearing in the middle of Georgian CC clusters do not have a lingual gesture. Furthermore, although CC timing is related with the occurrence of the vocoid, such that vocoids are more likely to be present when the timing is loose, the duration of the vocoid and the duration of the inter-consonantal lag are not correlated (Crouch et al. 2023b). This suggests that the transitional vocoid, even when present, is not an accurate acoustic correlate of the inter-consonantal timing, but its mere approximation. In French, on the other hand, a transitional vocoid has not been reported, presumably because the component consonants in an onset CC cluster are more tightly timed with each other (Bombien and Hoole 2013; Kühnert et al. 2006; Pouplier et al. 2022).

1.3.2 Stress and prominence

Georgian and French also systematically differ in stress and prominence patterns. Georgian has fixed word-initial stress in disyllabic and trisyllabic words (Borise and Zientarski 2018; Jun et al. 2007; Vicens and Jun 2014) and French has final prominence (e.g., Jun and Fougeron 2000; Vaissière 1983). The final prominence in French is known to be exclusively phrasal, rather than word-level, prominence.

In Georgian, word stress and phrasal stress are separate prosodic phenomena, characterized by different locations of prominence, and realized by different acoustic parameters. Georgian word-initial stress is realized by vowel duration and intensity as the main parameters, while phrasal prominence is cued by F0 targets at the right edge of the prosodic domain, on the antepenult and penultimate syllables. The most recent study known to us, Borise (2023), establishes duration as the main parameter of word stress in Georgian, and provides additional information on its complex interaction with phrasal prominence. Most relevant to our study is the finding that in disyllabic words, the vowel of the initial stressed syllable is longer than the vowel of the second syllable. It is especially relevant to note that, while unstressed syllables have shorter duration, unstressed vowels in Georgian are never reduced in their vowel quality.

French is a final prominence language, although the domain is larger than the word (AP – Accentual Phrase, Jun and Fougeron 2002, among others). If a word-final phonemic vowel occurs at an AP-final position, it is more prominent (longer in its duration – Adda-Decker et al. 2008; Meunier and Espesser 2011; and lower in its vowel height – Meunier and Espesser 2011) than non-final vowels. Meunier and Espesser (2011) show that in disyllabic words in a French corpus, /a/ in second syllables is longer and has higher F1 (lower in quality, open jaw) than the same vowels in the first syllables. While it cannot be concluded that the word-final vowels are always

prominent (because they might still be AP-internal), Meunier and Espesser (2011) suggest that word-internal vowels are always AP-internal, and thus are always reduced in terms of duration as well as in vowel quality.

The reduction of non-final, non-prominent vowels in French influences how the vowels are perceived by native listeners. Hall and Hume (2013) show that native listeners show low accuracy in identifying mid-front rounded vowels [œ] and [ø], and French “schwa” <e>, in [aCVCa] context. These three vowels are highly confusable with one another. Also, when there is no vowel (i.e., when the stimuli were [aCCa]), listeners in Hall and Hume (2013) report hearing a mid-front rounded vowel /ø/ for about 20 % of the time. Note that in the context used in Hall and Hume, the target vowel is in a non-final, non-prominent position. Similarly, Malécot and Chollet (1977) suggest that French “schwa” <e> and /ø/ are phonetically similar and that listeners cannot identify the vowels reliably.

To summarize, the two languages differ in their prominence and stress patterns (word initial stress in Georgian and accentual phrase-final prominence in French), as well as in the phonetic parameters associated with the prominence/stress patterns. The stressed or prominent syllables are longer than unstressed or non-prominent ones in both languages. The vowel quality reduction is reported only in French and the mid-front rounded vowel /ø/ becomes similar to schwa when it is not prominent and reduced.

1.4 The current study

This study aims to investigate the perception of illusory clusters by asking whether a nonnative CVCV sequence is perceptually repaired into a CCV sequence when the temporal organizations of the nonnative CVCV sequence and the native CCV sequence are similar. To explore this question, Georgian listeners are tested with French CCa-CVCá contrasts. Here, we present predictions based on PAM (e.g., Best 1995), which claims that the perceptual assimilation patterns will be determined by the articulatory similarities between the nonnative phones and the closest native counterparts.

Due to the reduction of the non-prominent V in French CVCá, together with the difference in the inter-consonantal timing between Georgian and French, the closest match of French CVCá, in terms of the temporal organization of the involved gestures, might not be Georgian CVCa but CCa. If the (mis-)match between the timing patterns in the listeners' native language and those in the stimuli can drive the perceptual assimilation patterns, both French CCa and CVCá would be assimilated to Georgian CCa, leading to inaccurate discrimination for the CCa-CVCá contrast by Georgian listeners. This is partly because V in French CVCá is reduced in its duration

and vowel quality relative to the prominent second vowel /a/, but more importantly, because Georgian /CCa/s can have relatively long inter-consonant lag that is often accompanied by a transitional vocoid between the two consonants.

We further predict that the probability of French CVCá being assimilated either to CCa or to CVCa by Georgian listeners depends on the similarity in the articulatory configuration between the two consonants, that is, during V in French CVCá and during the temporal void in Georgian CCa sequences. We consider three different V, /a/, /u/, and /ø/, among which French non-prominent /ø/ is known to be reduced to [ə] (e.g., Fougeron et al. 2007; Hall and Hume 2013; Meunier and Espesser 2011). The transitional vocoid in Georgian CC clusters, when present, is typically described as schwa (e.g., Chitoran et al. 2002; Crouch 2022). All of these would make the Georgian CCa a plausible assimilatory target for French CøCá sequences. Therefore, we predict that Georgian listeners are more likely to perceptually assimilate French CøCá, rather than CaCá or CuCá, to Georgian CCa, leading to less accurate discrimination of the CCa-CøCá contrast compared to the CCa-CaCá and CCa-CuCá contrasts.

This does not mean that Georgian CCa would be a perfect match for French CøCá. The absence of /ø/ in Georgian makes it vulnerable to perceptual repair, being assimilated to its closest “phone” in Georgian. The Georgian vowel inventory has 5 vowels /i, ε, α, ɔ, u/ (Robins and Waterson 1952; Shosted and Chikovani 2006), among which /u/ seems to be quite close to French /ø/, with lips being rounded and often fronted. Therefore, Georgian listeners are expected to assimilate French CøCá, which includes a nonnative phone /ø/, weakly to Georgian CCa repairing the nonnative phone /ø/ by an apparent segmental deletion, and perhaps more strongly to Georgian CúCa, if French /ø/ is phonetically similar to Georgian /u/.

These predictions are based on PAM (Best 1995) that explicitly indicates that the primitives of speech perception are articulatory gestures that have temporal as well as spatial dimensions. This, in our view, makes PAM the most appropriate theory to conceptualize the role of temporal structures beyond segments in nonnative speech perception in a straightforward way. However, we do not claim that PAM (or other theories based on articulatory gestures as the primitives) is the only theory that would predict the CVCV-to-CCV assimilation. For example, theories that view the perception of speech as a hypothesis testing process (e.g., Stevens and Halle’s 1967 Analysis-by-Synthesis model) or as a statistical inference (e.g., Feldman et al. 2009) would yield similar predictions via different mechanisms. As theories do not necessarily make competing predictions, we do not aim to assess different theories of speech perception. Instead, we aim to show, without denying the possibility of alternative explanations, how the CVCV-to-CCV assimilation can be conceptualized as the re-grouping of the involved gestures into segments that stems from the temporal organizations of the involved gestures.

2 Methods

2.1 Participants

Forty native speakers of Georgian were recruited at Tbilisi State University (Tbilisi, Georgia). All participants were adult native speakers of Georgian, but they were not monolinguals. Twenty-six Georgian participants reported knowing Russian to varying degrees, and thirty-two knowing English. Crucially, none of the participants reported knowing French. Data from four participants who reported learning a language with front rounded vowels were excluded. The languages included German (2), Azerbaijani (1), and Turkish (1). Data from three additional listeners were lost due to technical issues. After excluding the disqualified participants and lost data, data from thirty-three Georgian listeners were included in the analysis.

Forty-one Parisian French listeners were recruited at the Université Paris Cité (Paris, France) as the control group. All were adult native speakers of French, did not speak other languages on a regular basis, and had no prior experience of learning Georgian or other languages with a rich onset cluster inventory.

All participants gave their informed consent for participation in the study and for the subsequent use of their data. None reported any known history of speech or hearing impairments. They all received payment for their participation, in accordance with the rates used in the respective countries at the time of testing.

2.2 Stimuli

2.2.1 Preparation

The stimuli consisted of C(V)Cá pseudo-words including eight different CC combinations (/b/, /g/, /p/, /k/, /s/, /sk/, /ps/, and /pt/) and four different V conditions (/a/, /u/, /ø/, and 'no vowel'). This yielded four items per each CC combination (e.g., /balá/, /bulá/, /bølá/, and /bla/). All CC combinations are licit word-onset clusters in French, although /pt/ and /ps/ occur in a limited number of lexical items (Dell 1995). CC combinations with /ʁ/ (orthographic <r>) were intentionally excluded to avoid exposing Georgian listeners to an unfamiliar consonant. Since /ʁ/ does not have a clear counterpart in the Georgian inventory, it might present an additional challenge to Georgian listeners unrelated to testing how the nonnative timing is repaired.

A female native speaker of Parisian French recorded the 32 pseudo-words (8 CC combinations * 4 V conditions) in a set of carrier sentences: *Je {dis/lis/écris} ___ dans {le jardin/le salon/la cuisine}* 'I {say/read/write} ___ in the {garden/room/kitchen}'.

Each pseudo-word was repeated 4 times in randomized orders. Of the four repetitions, we selected two instances of each pseudo-word for inclusion according to the following criteria. First, tokens with any disturbance or deviant prosody were removed. Second, the selected two tokens of each pseudo-word had similar durations of the final vowel /á/. Third, only the tokens followed by a phrasal boundary (determined by the phrase-final pitch accent H*) were included. This was to make sure the first vowel in /CVCá/ was more reduced than the final /á/. The tokens were extracted from the carrier sentences from the point where the initial consonant was free from the coarticulatory information of the previous vowel to the F2 offset of the final vowel /á/. The selected tokens were then equalized to have an average intensity of 65 dB and concatenated to make the stimulus pairs, using Praat (Boersma and Weenink 2021).

For each consonant combination, 18 pairs (8 same pairs, 10 different pairs) were created. The “same” pairs included two tokens of each pseudo-word that were not acoustically identical but phonologically (or lexically) equivalent to the French speaker who produced the stimuli. For example, the eight same pairs for the consonant combination /bl/ included /bla_A-/bla_B, /balá_A-/balá_B, /bulá_A-/bulá_B, /bølá_A-/bølá_B, and their mirror images. The “different” pairs included six in which the ‘no vowel’ tokens (i.e., CCa sequences) were paired with CVCá sequences (/bla_A-/balá_A, /bla_A-/bulá_A, /bla_A-/bølá_A, and their mirror images) as well as four pairs containing two CVCá sequences (/bala_A-/bølá_A, /bula_A-/bølá_A, and their mirror images). This yielded a total of 144 distinct pairs: 18 pairs (8 same, 10 different) * 8 CC combinations.

2.2.2 Acoustic analysis

Table 1 presents the means and standard deviations of the duration and formant measurements of the V in the CVCá stimuli (see Supplementary Material for the measurements of individual tokens). Formant measures were taken from the temporal midpoint of the vowels. Duration ratio was calculated using the following formula: $\frac{\text{duration of V in CVCá}}{\text{duration of /á/ in CVCá}} \times 100$ (%). The acoustic measures revealed some interesting observations relevant to the current investigation. First, the duration ratio indicates that the first vowel was shorter than the final /á/, confirming that French stimuli were produced with final prominence. Second, the formant measures suggest that French /ø/ was indeed centralized (i.e., schwa-like), as previously reported (e.g., Hall and Hume 2013). This would make French /ø/ quite similar to Georgian transitional vocoids in terms of tongue configuration (and vowel quality).

As expected, all but one French stimulus did not include a transitional vocoid between the two consonants in CCV stimuli. When the two component consonants were voiceless, no CCV tokens included a vocoid with a glottal pulsing in the

Table 1: Mean acoustic measurements of V in CVCá tokens in the French stimuli (standard deviation in parenthesis).

Vowel	Duration (ms)	Duration ratio (%)	F1 (Hz)	F2 (Hz)
a	69.3 (12.0)	93.5 (18.3)	646 (32)	1878 (167)
ø	61.2 (15.7)	77.4 (22.5)	417 (22)	1595 (119)
u	54.9 (17.6)	66.2 (22.5)	326 (50)	1175 (157)

waveform or a voicing bar in the spectrogram. Among the tokens including one or more voiced component consonant, only one token of /bla/ had a vocalic element with a relatively greater amplitude, distinctive from /l/ in the waveform and the spectrogram (see Figure 1). The vocoid was 35.2 ms long, shorter than the phonemic vowels in CVCá tokens (Table 1). F1 and F2 of this vocoid were 383 Hz and 1968 Hz, respectively, which would make the quality of this vocoid comparable to a mid-high (slightly lower than /u/) central (slightly fronter than /a/) vowel.

2.3 Procedure and task

The experiment consisted of a same-different discrimination task, in which the “same” trials included two phonologically equivalent but acoustically different tokens. The listeners were seated in front of a MacBook Pro laptop, with a response pad (model RB-740, Cedrus Corporation) attached. On each trial, the participants heard a pair of

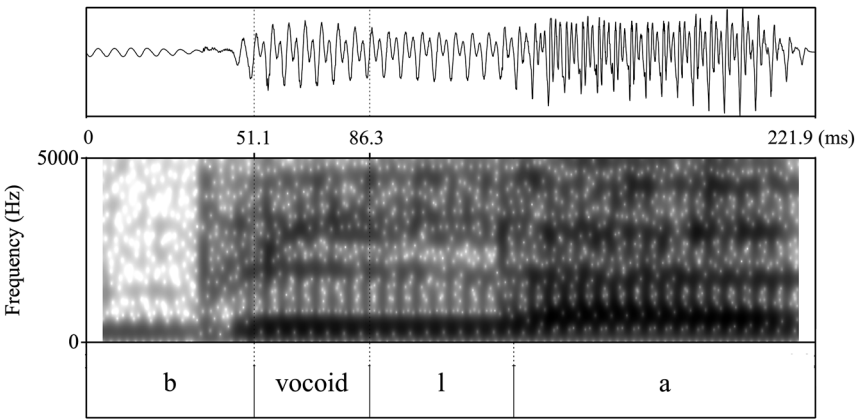


Figure 1: The sound waves (top) and the spectrogram (bottom) of a token of /bla/. The vocoid is separated from /l/ in its amplitude and formant.

“words” over headphones (AKG K271 MK II) and were instructed to determine whether they heard two different “words” or two repetitions of one “word”. They responded by hitting one of the two designated buttons on the response pad. The two buttons were marked with initials for “same” and “different” in the listeners’ native languages (e.g., Georgian: <ი> for იგივე “same”, <გ> for განსხვავებული “different”; French: <M> for *même* “same”, <D> for *différent* “different”). The task was self-paced, and each new trial played 1000 ms after the participant hit the button for the previous trial. All stimulus presentation and data collection were implemented using PsychoPy2 (version 1.85.2, Peirce et al. 2019). Listeners were told that the stimuli may include a foreign language, but they were not informed which language it would be.

Participants were first provided with 8 practice trials to familiarize themselves with the task. Half of the practice trials were “same” and the other half were “different”. The practice trials were structurally similar to the test trials (i.e., involving CCa and CVCa), but included different tokens. After the practice trials, participants had a chance to ask the experimenter any questions they had, after which the main experiment started. During the main experiment, Georgian listeners completed two blocks separated by a self-terminated break. Each block presented one repetition of the entire stimuli ($n = 144$) in randomized orders. The control group (French listeners) completed only one block.

All written instructions during the experiment, including the survey and the consent forms, were provided in the listeners’ native languages. For oral communications, a bilingual speaker of Georgian and English helped the experimenters interact with the Georgian participants in Georgian. The experimenters interacted with the French participants in French.

All participants were tested in an additional perception experiment, either before or after the current experiment, and a separate production study after the perception experiments. The additional perception experiment was similar to the current one but used stimuli in a different language and the production study involved producing CCV and CVCV tokens. After completing all the procedures, participants completed a self-report language background survey.

2.4 Analysis

Prior to the analysis, we removed the responses with their response times not within 2.5 standard deviations of the mean response time for each participant (379 out of 15,408 responses). The remaining responses (same or different) were converted to a sensitivity measure d_a , based on the principles of Signal Detection Theory (Macmillan and Creelman 2005), using the following formula: $d_a = [2/(1 + s^2)]^{1/2} \times [z(\text{hit rate}) - sz(\text{false alarm rate})]$. In the formula, s refers to the ratio of same (noise) to

different (signal) distributions. This measure of sensitivity is deemed more appropriate than a more commonly used measure d' , when the variances of signal and noise are expected to be unequal (e.g., Simpson and Fitter 1973; Verde et al. 2006).

As we aimed to compare the listeners' sensitivity in the five examined contrasts (CCá-CaCá, CCá-CuCá, CCá-CøCá, CøCá-CaCá, and CøCá-CuCá), d_a values were calculated separately for each listener and for each of the five contrasts. That is, each d_a value was based on one listener's responses on six distinct trials for each of the eight CC combinations, four of which were same trials and two were different. Table 2 demonstrates the six distinct trials for each of the five contrasts when the CC combination was /bl/. Combining the five contrasts, each CC combination included eight unique same trials (the "same" column in Table 2) and 10 different trials (the "different" column in Table 2). For the s in the d_a formula, we used 2, the ratio of same to different trials for each contrast. Twice as many same trials as different trials were included in d_a calculation for each contrast as we wanted the listeners to experience not too many different trials compared to the same trials during the task. In other words, we doubled the number of the same trials to decrease the size of a potential response bias (Macmillan and Creelman 2005). The same-to-different ratio that the listeners experienced during the task was 4:5 (see Section 2.2.1), which would have been 2:5 without doubling the number of the same trials.

The extreme values for the hit and false alarm rates were corrected using the log-linear methods in Hautus (1995). Since French listeners, the control group, heard only one repetition of the stimuli while Georgians heard two repetitions, French listeners' number of trials were multiplied by two before applying the log-linear correction. This was to prevent the size of distortion caused by the log-linear correction from being different from Georgian listeners to French controls. A Georgian listener had five d_a values (one per contrast), each comprising 96 same/different responses (6 trials * 8 CC combinations * 2 repetitions). In the case of French listeners, each d_a value was based on 48 responses (6 trials * 8 CC combinations).

Table 2: Trials used in d_a calculation when CC combination was /bl/.

Contrast	Same	Different
CCa-CaCá	/bla/_A-/bla/_B, /bla/_B-/bla/_A, /balá/_A-/balá/_B, /balá/_B-/balá/_A	/bla/_A-/balá/_A, /balá/_A-/bla/_A
CCa-CuCá	/bla/_A-/bla/_B, /bla/_B-/bla/_A, /bulá/_A-/bulá/_B, /bulá/_B-/bulá/_A	/bla/_A-/bulá/_A, /bulá/_A-/bla/_A
CCa-CøCá	/bla/_A-/bla/_B, /bla/_B-/bla/_A, /bølá/_A-/bølá/_B, /bølá/_B-/bølá/_A	/bla/_A-/bølá/_A, /bølá/_A-/bla/_A
CøCá-CaCá	/bølá/_A-/bølá/_B, /bølá/_B-/bølá/_A, /balá/_A-/balá/_B, /balá/_B-/balá/_A	/bølá/_A-/balá/_A, /balá/_A-/bølá/_A
CøCá-CuCá	/bølá/_A-/bølá/_B, /bølá/_B-/bølá/_A, /bulá/_A-/bulá/_B, /bulá/_B-/bulá/_A	/bølá/_A-/bulá/_A, /bulá/_A-/bølá/_A

3 Results

Figure 2 shows the d_a scores of Georgian listeners along with those of the French controls. For all five contrasts, Georgian listeners seem to have lower sensitivity than French controls to varying extents. To examine for which contrast(s) Georgian listeners' sensitivity differed from French listeners, the d_a scores were statistically analyzed by building a series of linear mixed effects models, using the *lme4* package (Bates et al. 2015) in R (R Core Team 2021). We first built the full model with CONTRAST (CCa-CaCá, CCa-CuCá, CCa-CøCá, CøCá-CaCá, CøCá-CuCá) and listeners' native LANGUAGE (Georgian, French) as the fixed factors, along with their interactions. CONTRAST was Helmert-coded while LANGUAGE was dummy coded with the reference level being

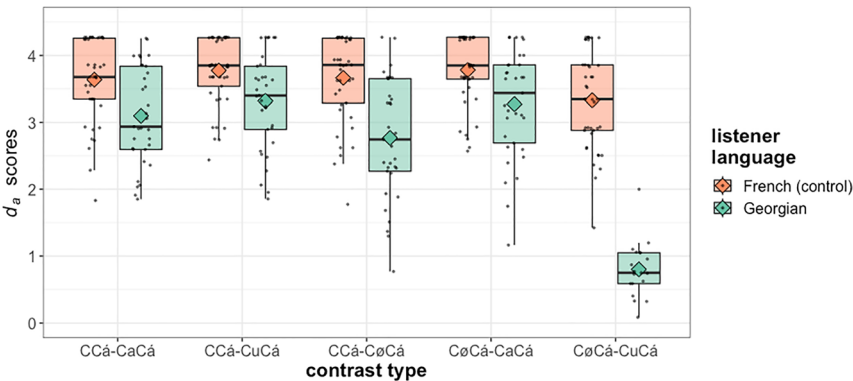


Figure 2: Listeners' sensitivity (d_a scores) to five French contrasts. Diamonds represent the mean values.

Table 3: Model outcome.

Fixed effects	Estimate (β)	Standard error	t values
Intercept	3.638	0.083	43.869
Language (Georgian)	-1.102	0.124	-8.878
Contrast1 (CCa-CuCá)	0.068	0.057	1.209
Contrast2 (CCa-CøCá)	-0.014	0.033	-0.415
Contrast3 (CøCá-CaCá)	0.023	0.023	0.978
Contrast4 (CøCá-CuCá)	-0.077	0.018	-4.287
Language: Contrast1	0.042	0.085	0.495
Language: Contrast2	-0.133	0.049	-2.723
Language: Contrast3	0.030	0.035	0.871
Language: Contrast4	-0.503	0.027	-18.786

set to French (the control group). For the random effects, by-SUBJECT intercept was included. Adding random slopes led to a singular fit or a convergence error. The outcome of the full model is in Table 3.

This full model was compared with the model without the LANGUAGE: CONTRAST interaction using a likelihood ratio test, which revealed that the interaction was significant [$\chi^2(4) = 240.7, p < 0.001$]. This significant interaction was further

Table 4: Pairwise comparisons for listeners’ native language.

Language	Contrast	Estimate (β) ^a	<i>p</i> values ^b
French – Georgian	CCá-CaCá	0.538	0.001
	CCá-CuCá	0.454	0.006
	CCá-CøCá	0.896	<0.001
	CøCá-CaCá	0.509	0.002
	CøCá-CuCá	3.115	<0.001

^aPositive β values indicate greater sensitivity in French listeners than in Georgian listeners. ^b*p* values were adjusted using the Tukey method.

Table 5: Pairwise comparisons for contrast.

Language	Contrast	Estimate (β) ^a	<i>p</i> values ^b
Georgian	CCá-CaCá – CCá-CuCá	−0.221	0.405
	CCá-CaCá – CCá-CøCá	0.330	0.070
	CCá-CaCá – CøCá-CaCá	−0.174	0.639
	CCá-CaCá – CøCá-CuCá	2.883	<0.001
	CCá-CuCá – CCá-CøCá	0.550	<0.001
	CCá-CuCá – CøCá-CaCá	0.046	0.996
	CCá-CuCá – CøCá-CuCá	3.104	<0.001
	CCá-CøCá – CøCá-CaCá	−0.504	0.001
	CCá-CøCá – CøCá-CuCá	2.553	<0.001
	CøCá-CaCá – CøCá-CuCá	3.058	<0.001
French	CCá-CaCá – CCá-CuCá	−0.137	0.746
	CCá-CaCá – CCá-CøCá	−0.028	0.999
	CCá-CaCá – CøCá-CaCá	−0.145	0.702
	CCá-CaCá – CøCá-CuCá	0.306	0.056
	CCá-CuCá – CCá-CøCá	0.109	0.871
	CCá-CuCá – CøCá-CaCá	−0.008	1.000
	CCá-CuCá – CøCá-CuCá	0.443	0.001
	CCá-CøCá – CøCá-CaCá	−0.117	0.838
	CCá-CøCá – CøCá-CuCá	0.334	0.028
	CøCá-CaCá – CøCá-CuCá	0.451	0.001

^aPositive β values indicate greater sensitivity in the first pair than the second pair in the Contrast column. ^b*p* values were adjusted using the Tukey method.

examined, without attempting to test the main effects of LANGUAGE or CONTRAST, with post-hoc pairwise comparisons in the *emmeans* package (Lenth 2020). The *p*-values for the pairwise comparisons were adjusted using the Tukey method. The results of these pairwise comparisons are summarized in Tables 4 and 5.

Throughout all the contrasts, Georgian listeners showed significantly lower sensitivity than French listeners (Table 4). This was the case even when the stimuli did not include the nonnative phone [ø], such as in the contrast CCá-CaCá or CCá-CuCá. The pairwise comparisons in Table 5 suggest both Georgian and French listeners' sensitivity was influenced by CONTRAST, as also shown in Figure 2. Related to our question are the contrasts including [ø], namely the CCa-CøCá, CøCá-CaCá, and CøCá-CuCá contrasts. For Georgian listeners, discrimination of the CøCá-CuCá contrast was significantly less accurate than that of the other contrasts [all *p*'s < 0.001], as expected. But CøCá-CuCá was not the only contrast that Georgian listeners had difficulty with. Their discrimination of the CCa-CøCá contrast was also less accurate than the CCa-CuCá contrast [$|\beta| = 0.550$, $p < 0.001$] and than CøCá-CaCá [$|\beta| = 0.504$, $p < 0.001$]. The difference between CCa-CøCá and CCa-CaCá contrasts did not reach significance [$p = 0.070$]. This cautiously suggests that Georgians' sensitivity to the CCa-CaCá contrast may also have been somewhat low compared to the CCa-CuCá/CøCá-CaCá contrasts whose d_a scores were significantly greater than those of the CCa-CøCá contrast. The outcomes suggest that Georgian listeners, as predicted, had difficulty with the nonnative phone [ø]. It is particularly intriguing that Georgian listeners confused French CøCá not only with French CuCá but also with French CCa, though not as frequently. And, to an even smaller extent, Georgian listeners may have confused French CaCá with French CCa.

The CøCá-CuCá contrast also showed lower discrimination accuracy in French listeners (the control group) compared to some of the other contrasts (e.g., CøCá-CaCá, CCá-CøCá, CCa-CuCá). The difference between the CøCá-CuCá contrast and the CCa-CaCá contrast was marginally significant [$p = 0.056$]. This presumably suggests that the French front rounded vowel /ø/ and back rounded vowel /u/ bear some phonetic similarities, causing them to be confused even by the native listeners. Crucially, the current outcome did not provide statistical evidence that French listeners had lower sensitivity for the CCa-CøCá contrast. As shown in Table 5, French listeners' sensitivity for the CCa-CøCá contrast did not differ significantly from those for the CCa-CaCá, CCa-CuCá, or CøCá-CaCá contrasts, and was significantly greater than those for the CøCá-CuCá contrast. This starkly contrasts with the Georgian listeners' case.

To further understand the source of Georgian listeners' low sensitivity to the CCa-CøCá contrast, we further examined the CCa-CøCá pairs, breaking them down according to the consonant combinations. Table 6 presents the mean d_a scores of Georgian listeners on the French CCa-CøCá contrast for each CC combination with

Table 6: Georgian listeners’ sensitivity to CCa-CøCá contrast by CC composition.

CC	Mean d_a [95 % CI]	Significantly ^a higher than	Significantly lower than
sk	2.41 [2.17–2.65]	gl, pt, bl, pl, sp	
ps	2.35 [2.11–2.59]	gl, pt, bl, pl, sp	
kl	2.24 [1.99–2.48]	pt, bl, pl, sp	
gl	1.86 [1.61–2.10]	bl, pl, sp	sk, ps
pt	1.56 [1.32–1.80]	pl, sp	sk, ps, kl
bl	1.36 [1.12–1.61]	pl, sp	sk, ps, kl, gl
pl	–1.01 [–1.25 to –0.77]		sk, ps, kl, gl, pt, bl
sp	–1.01 [–1.25 to –0.77]		sk, ps, kl, gl, pt, bl

^aTukey adjusted $p < 0.05$ in pairwise comparisons conducted using *emmeans()*.

their 95 % confidence intervals. The sensitivity data suggest that the composition of the consonant clusters indeed influenced the discriminability. An additional linear mixed effects model was fitted to the CCa-CøCá d_a data, with the fixed factor of CC composition and the random intercept for participants. This model, when compared to the model without the CC composition in a likelihood ratio test, confirmed a significant effect of CC [$\chi^2(7) = 426.2, p < 0.001$]. This suggests that Georgian listeners did not have difficulty with all CCa-CøCá pairs to equal extents. The last two columns in Table 6 show which CC combinations significantly differ from one another, determined by the post-hoc pairwise comparisons implemented in *emmeans()*. It is not straightforward to attribute the different sensitivity scores to a specific consonant as the C₁ or C₂. For example, /sk/ had the highest and /sp/ had the lowest d_a scores though both clusters are sibilant-initial. The clusters including /l/ as C₂ also showed a wide range of sensitivity, with the velar-/l/ clusters showing higher sensitivity scores than the labial-/l/ clusters.

The same comparison on French listeners’ d_a scores of the CCa-CøCá pairs did not reveal a significant effect of CC composition [$\chi^2(7) = 5.004, p = 0.67$]. While the lack of the CC composition effect may be due to the ceiling effect (Figure 2), it still suggests that French listeners’ discrimination of the CCa-CøCá pairs was not influenced by the CC composition in the same way as Georgian listeners. This arguably precludes the possibility that the CC composition effect observed in Georgian listeners’ CCa-CøCá discrimination is simply due to the characteristics of the stimuli (see more on this in Section 4.2).

4 Discussion

In this study, we tested Georgian listeners’ discrimination of French CCa-CVCá pairs, aiming to examine how temporal organization of CC clusters (both in the nonnative

speech signals and in the prevalent or typical patterns in the listeners' language) may influence the perceptual repair patterns. The results revealed that French CVCá sequences with the nonnative vowel /ø/ were not exclusively confused with those with the vowel /u/, but also with CCa sequences without a phonemic vowel between the two consonants, albeit to a smaller extent. Note that only one French CCá token /bla/ included an apparent transitional vocoid between the two component consonants (Figure 1), and Georgian word-initial CC clusters have a greater variation in inter-consonantal timing, ranging from short to long lags (Poupplier et al. 2022). Therefore, it is highly unlikely that the Georgian listeners would have assimilated French CCa sequences to Georgian CVCa sequences. This indicates that Georgian listeners assimilated French /ø/ in CøCá to the temporal void resulting from the transition between the two consonants in Georgian CCa sequences. The typical temporal implementation of Georgian word-initial CC clusters seems to have influenced Georgian listeners' discrimination of CøCá and CCá pairs, which may further suggest that the temporal organizations of onset CC clusters are language-specific and thus should be included in the phonetic grammar, as expanded in Section 4.2.

In Section 4.1, these findings will be discussed with respect to the taxonomy of assimilation patterns in PAM (e.g., Best 1995). Then we discuss the implications of our findings for this theory of nonnative speech perception in Section 4.2.

4.1 Interpretation of the findings

According to PAM (e.g., Best 1995), nonnative phones are perceptually assimilated to the closest native phones, and the patterns of this perceptual assimilation are determined by the articulatory similarities (or discrepancies) between the nonnative and native phones. This assimilation pattern (i.e., how the members of a contrast are assimilated to the native categories), in turn, predicts the discrimination of a nonnative contrast.

Georgian listeners' poor discrimination between French CøCá-CuCá (see Figure 2) suggests that the contrast is assimilated as SC (Single Category), confirming the prediction that French /ø/ would likely be perceived as a reasonably good exemplar of /u/ to Georgian listeners. Georgian listeners presumably assimilate both French CøCá and French CuCá to Georgian CúCa, arguably with a small category-goodness difference, leading to poor discrimination between the two.

More relevant to our discussion is the CCa-CøCá contrast, whose discrimination is not nearly as bad as the CøCá-CuCá contrast, but still worse than CøCá-CaCá and CCa-CuCá, in contrast to the native controls (Figure 2, Table 5). This difference between CCa-CøCá and the French contrasts that are undoubtedly assimilated as TC (CøCá-CaCá, CCa-CuCá) is statistically significant although small in its magnitude

(Table 5), suggesting that the contrast is presumably assimilated as CG – both sequences (CøCá and CCa) are assimilated to a single native sequence with a relatively large category-goodness difference. This indicates that, when confronted with the sequences including a nonnative phone /ø/, Georgian listeners assimilate the input to the closest “phone” in their native language in more than one way. French CøCá sequences containing the nonnative vowel were predominantly assimilated to CúCa in Georgian, suggested by the poor discrimination between French CøCá and CuCá. This is a case of perceptual repair in terms of the vowel place of articulation (ø-to-u) and the prominence pattern (CVCV̆-to-CVCV̆). To a smaller extent, however, French CøCá sequences were confused with CCa. This outcome indicates that Georgian listeners perceptually assimilate French CøCá sequences to Georgian CCa. The closest “phone” to the nonnative phone /ø/, in this case, would be the open vocal tract between the consonants. We claim that this CVCV̆-to-CCV̆ assimilation can be explained as the re-parsing of the involved gestures into segments according to the dominant pattern of inter-gestural timing in the listeners’ native language (more on this in Section 4.2).

We acknowledge that the current findings alone do not provide definitive answers to whether CCa and CøCá are commonly assimilated to Georgian CCa or to something else (such as CúCa). Still, we argue it is highly unlikely for the French CCa stimuli to be assimilated to Georgian CVCá, regardless of the quality of the first vowel, since French CCa stimuli were produced with a quick transition between the two component consonants as they typically are (Bombien and Hoole 2013), Georgian allows a greater variation in the inter-consonantal timing patterns (Poupplier et al. 2022), and Georgian CVCá has more prominent and longer first V̆ than the final /a/ (Borise 2023). As mentioned in Section 2.2.2, French CCa stimuli used in this study did not have a transitional vocoid except for one token of /bla/. On the other hand, it is quite plausible that French CøCá, with the first non-prominent vowel having a schwa-like quality (Table 1), can be perceived by Georgian listeners as an exemplar of Georgian CCa, albeit not an ideal one. Therefore, we argue that the common assimilatory target for French CCa and CøCá is Georgian CCa.

Georgian listeners’ poor discrimination of French CCa and CøCá, then, suggests that the listeners know (as part of their language-specific phonetic knowledge) that a consonant cluster can be produced with quite a long lag between consonants. Consequently, they perceive French unaccented /ø/, reduced both in its vowel quality and in its duration in the context of CøCá, as a part of the consonant cluster. A similar case has been reported in Berent et al. (2009), as mentioned in Section 1.1. Russian listeners in Berent et al. (2009) sometimes (mis-)perceive English [CəCVC] beginning with a nasal consonant as Russian /CCVC/, and the authors claim that Russian listeners perceptually modified an unfamiliar structure (pretonic schwa) to a more acceptable structure (NC consonant cluster) in Russian. We suspect that this may also

be, at least partially, attributable to the different timing patterns in Russian and English. Hypothetically, if listeners of a language that only allows a tight timing between component consonants in onset CC clusters were tested with English [CəCVC], they would not assimilate it to /CCV/ even if their language does not allow pretonic schwas. We leave this for a future investigation.

It should also be noted that the low sensitivity in the current findings does not suggest that the listeners are “deaf” to the acoustic differences. The sensitivity measures averaging the CC combinations are above chance-level for all tested contrasts ($d_a > 0$, see Figure 2). More importantly, the listeners were not asked to determine whether the two tokens within a pair are acoustically identical. Instead, they were asked to judge whether the two acoustically different tokens within a pair were instances of the same word or two different words. When Georgian listeners responded that French CCa and CəCá were the same, for instance, it would have not been the case that they perceived the two tokens as being identical. Rather, they presumably detected some acoustic differences between the two tokens, judged the detected differences to be linguistically irrelevant, and thus “ignored” the differences.

4.2 Theoretical implications

Our findings raise an interesting question to PAM (Best 1995) as to what exactly counts as the native categories or phones to which the nonnative sounds can be assimilated. When Georgian listeners assimilate French CəCá to Georgian CCa, what is the Georgian phone that is determined to be the most similar to French /ə/? We claim that the closest phone to the nonnative phone /ə/, in this case, would be the temporal void rising from the timing pattern between the two consonants rather than the transitional vocoid that may (or may not) appear as an acoustic artifact of the temporal void. Articulatorily, the temporal void can be characterized as an open oral cavity between two component consonants. While it may not count as an active lingual gesture (Browman and Goldstein 1992; Crouch et al. 2023b), it directly results from the temporal relation between consonantal gestures. We propose that PAM should include the temporal organization in-between gestures, such as the one resulting in the temporal void here, as a possible assimilatory target. It is the temporal organization behind the articulatory event (i.e., open oral cavity) that can make this non-gestural articulatory event the target of assimilation in nonnative speech perception, as we argue below.

The temporal organization of word-initial CC clusters varies both within and across languages. Cross-linguistically, the variation in timing is not random but predictable from the composition of the consonants (Chitoran et al. 2002; Crouch

et al. 2023a; Hoole et al. 2009; Pouplier et al. 2022), indicating that the temporal organization cannot be reduced to the mere (bio-)mechanics of the vocal tract. Speakers seem to organize consonant gestures with more temporal distance, when tighter timing could result in unfavorable perceptual consequences, such as in back-to-front clusters than front-to-back clusters (Chitoran et al. 2002), or stop-nasal clusters than in stop-oral clusters (Hoole et al. 2009). On the contrary, tighter timing seems to be tolerated when it would not work against the perceptual recoverability of the consonants (e.g., sibilant-initial clusters, Pouplier et al. 2022). Also, specifically in Georgian, sonority sequencing of the onset clusters is systematically correlated with the inter-consonantal lags such that differences in timing facilitate lexical recoverability (Crouch et al. 2023a). In addition, all labial-velar or coronal-velar clusters in Georgian agree in their laryngeal specifications (“harmonic” clusters, see more details in Chitoran et al. 2002), suggesting that the tighter timing in front-to-back clusters, though it might have been perceptually motivated, may have phonological implications in contemporary Georgian. All these aspects point to the interpretation that the temporal organizations between component consonants within a cluster constitute language-specific knowledge. That is, speakers know, as language-specific phonetic knowledge, how onset clusters are typically timed in their native language.

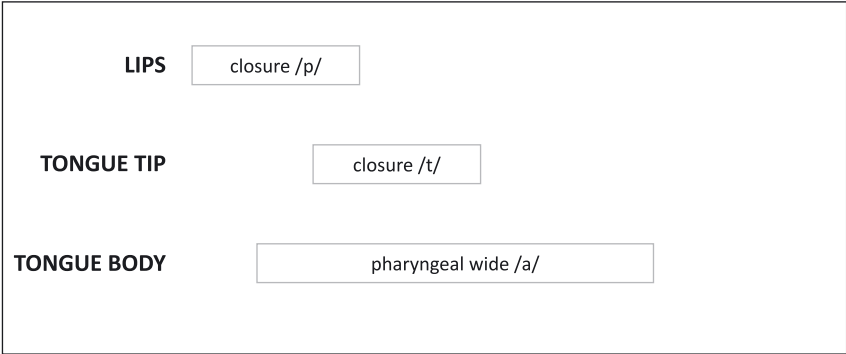
When the component consonants are loosely timed with a long inter-consonantal lag, the temporal void in-between results in, articulatorily, an open oral cavity between the consonants, and acoustically, a transitional vocoid (e.g., Chitoran et al. 2002; Davidson 2005). In terms of articulation, this temporal void is not associated with a specific lingual gesture (Crouch et al. 2023b), nor does it count as an active gesture (Browman and Goldstein 1992). Acoustically, the transitional vocoid occurs quite often, but not always, when the inter-consonantal lag is long (e.g., Chitoran et al. 2002; Crouch et al. 2023b). At the same time, the vocoid is almost systematically missing when both consonants in CC are voiceless (e.g., Chitoran et al. 2002; Crouch 2022; Pouplier et al. 2020). These suggest that the vocoid, as well as the open oral cavity that gives rise to the vocoid when the flanking consonants are voiced, is an artifact of the gestural timing. In other words, the language-specific knowledge does not likely specify the existence of the open oral cavity or the transitional vocoid within certain CC clusters. Rather, the speakers know the temporal organization of the involved gestures that gives rise to the open vocal tract and, in combination with other factors such as voicing of the consonants, the transitional vocoid.

We claim that this phonetic knowledge on language-specific inter-consonantal timing is the impetus for the perception of illusory clusters. That is, using the terminology of gesture-based theories (e.g., Best 1995; Browman and Goldstein 1992; Fowler 1996), Georgian listeners have the phonetic knowledge about how consonants

within an onset cluster are typically timed with one another, and this knowledge determines how the perceived gestures would be re-constellated into segments. The process of this re-constellation is illustrated in Figures 3 and 4, with the schematic gestural scores for /pta/ and /pVta/ sequences in French and Georgian. Figures 3 and 4 are only for illustrative purposes, and the explanation applies not only for the consonant sequence /pt/ but for other sequences as well.

Georgian /pta/, as shown in Figure 4(a), has a long timing lag between the lip gesture for /p/ and tongue tip gesture for /t/. When a Georgian listener hears French /pøta/, they have two competing candidates for the assimilatory targets, /puta/ and /pta/. French /pøta/, with the lip rounding gesture and voicing for /ø/, is close to Georgian /puta/ in terms of the spatial similarities among the involved gestures, as

(a) French /pta/



(b) French /pøta/

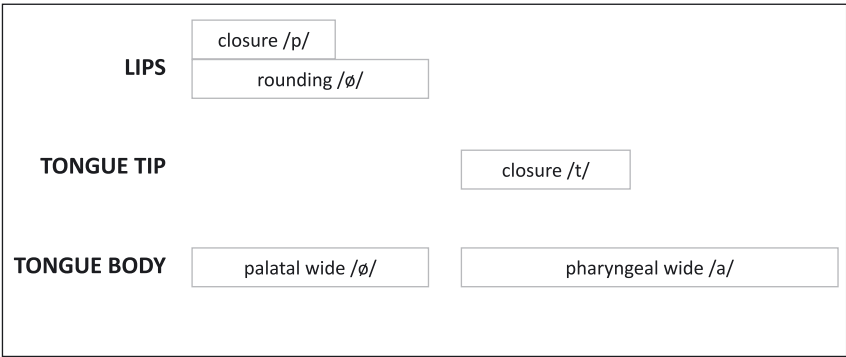


Figure 3: Schematic gestural scores for French (a) /pta/ and (b) /pøta/. Glottis and velum are omitted for simplicity.

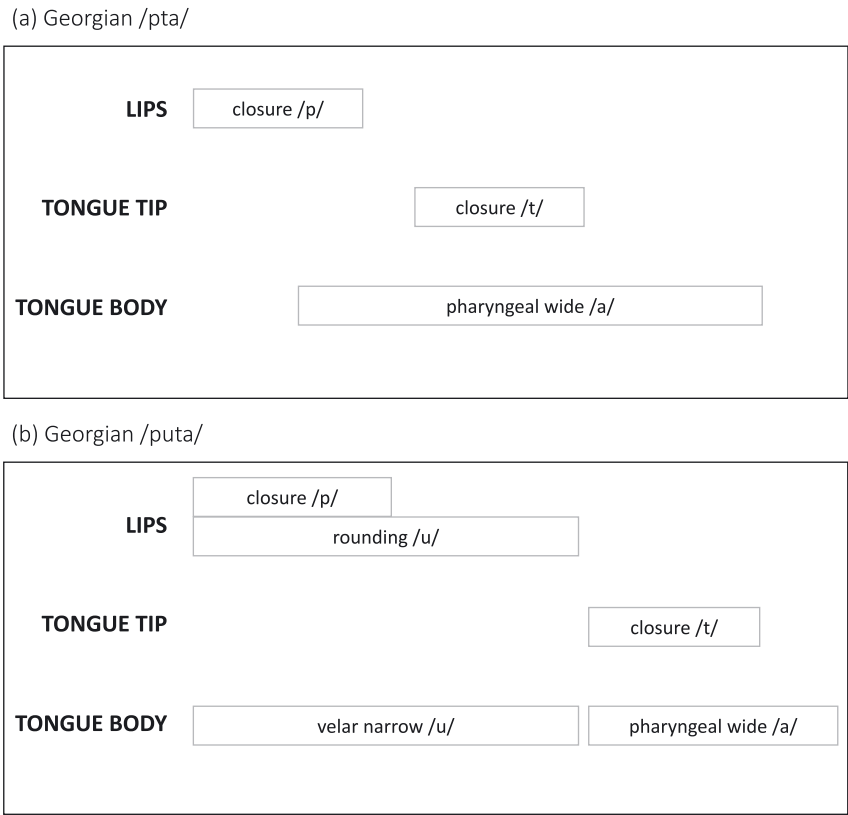


Figure 4: Schematic gestural scores for Georgian (a) /pta/ and (b) /puta/. Glottis and velum are omitted for simplicity.

shown in Figure 3(b) and Figure 4(b). The only spatial difference is from the location of the highest position of the tongue body, fronter in French /ø/ than in Georgian /u/. However, if the temporal organization is taken into consideration, French /pøta/ may be quite similar to Georgian /pta/ (Figure 3(b) and Figure 4(a)), especially when the non-prominent /ø/ is reduced (i.e., produced with a weak, if any, lip rounding gesture and even shorter duration). That is, a Georgian listener assimilating French /pøta/ to /pta/ can be explained by the similarities in the timing between the gestures involved.

Georgian listeners also had some trouble with discriminating French CCa and CaCá sequences. The sensitivity of the CCa-CaCá pairs was only marginally greater than that of CCa-CøCá pairs, unlike the CCa-CuCá/CøCá-CaCá pairs that showed highly accurate discrimination (see Table 5). This suggests that the unstressed /a/ in the CaCá context was also assimilated to the temporal void in Georgian CCa, albeit to a smaller

extent than the nonnative phone / \emptyset /. In search of the closest match of French /a/ in Georgian native categories, Georgian /a/ would be considered to be the best fit as they have similar tongue shapes. Still, because of the prosodic difference between the two languages, when the temporal aspects of the involved gestures (i.e., duration of the gestures and the phasing relations) are taken into account, the Georgian category that is closest to French unstressed /a/ in this specific context (CaCá sequence) is no longer the Georgian vowel /a/, but the temporal gap between the consonants. The temporal proximity seems to play a role, though small, even when the nonnative phone has a very close match among the native vowels in terms of the spatial properties of the involved gestures.

The perceptual assimilation patterns described above refer not only to the spatial information about the gestures (i.e., the configuration of the tongue or the lips) but also to their temporal organizations. Temporal perceptual repair beyond segmental boundaries has earlier been claimed by Best and Hallé (2010) who showed that the Zulu lateral fricative was often perceived as a consonant cluster by English and French listeners. Two simultaneous gestures in the Zulu lateral fricative were perceived as sequential, which is more consistent with the pattern in the listeners' language. Best and Hallé's (2010) findings suggest that the constellations of specifically-timed gestures can act as the native categories with the concept of segments not necessarily involved. Gestures can be re-constellated to match, as closely as possible, the typical gesture-segment mappings in the listeners' native language. Best and Hallé (2010) show gestural re-constellation within the onset structure involving one or more segments (singleton consonants, consonant clusters, or affricates), but our findings extend their findings to sequences of segments even across a syllable boundary. Nonnative speech perception does not simply involve segment-to-segment mapping. Rather, the process of perceptual assimilation simultaneously considers an array of factors which include not only the involved articulators and their constriction properties, but also their temporal organization. And when the segmental or syllabic affiliation of the involved gestures and the temporal organizations provide mismatching information in terms of the listeners' language, the temporal information can sometimes win the game.

Crucially, the association between segments and articulatory events is language-specific. When listeners hear an unfamiliar language with no prior exposure, they will not have the knowledge about this gesture-to-segment association, or mapping, in the stimulus language. Therefore, when the stimulus language and the listener's language(s) differ in the mapping, listeners would resort to the mappings that they are familiar with (i.e., the typical patterns in their native language(s)). For example, an open oral cavity with a certain duration between two consonants must belong to a vowel in French, but not necessarily in Georgian. The open oral cavity between the two consonants in French /p \emptyset ta/ may be perceived by Georgian listeners as a vowel

(in Georgian /puta/) or as the temporal void between the component consonants within the cluster (Georgian /pta/). Regardless of its segmental affiliation in the stimulus language, an articulatory event would be perceived based on the listeners' phonetic knowledge about how it is typically timed relative to another articulatory event and how it is typically associated with a segment in their native language.

Though this provides a simple explanation to Georgian listeners' perception of illusory clusters without additional complications, we do not claim that our findings provide unequivocal support to PAM (or other phonetic theories that take articulatory gestures as the primitives). Acoustically speaking, the assimilation patterns explained above suggest that nonnative speech perception needs to take into account sub-phonemic phonetic details and listeners may not access the syllabic or segmental affiliations of certain phonetic properties in an unfamiliar nonnative language. Also, in assimilating nonnative vocalic sounds, the acoustic correlates of the lingual articulation (i.e., formants) need to be considered simultaneously with the acoustic correlate of the temporal organization. And as the transitional vocoid is not an accurate acoustic correlate of the temporal organization of the CC cluster (Crouch et al. 2023b), listeners would need to attend to temporal relations among other acoustic cues that provide information on the closures or the releases of the flanking consonants.

Finally, the outcomes of the CC composition effects on the sensitivity to CCa-CøCá pairs (Table 6) are surprising when considering the previous findings on articulatory timing of consonant clusters. Sibilant-initial clusters are reported to be tightly timed cross-linguistically (e.g., Pouplier et al. 2022) so /sCa/ would have a distinct temporal organization from /sVCá/. This makes the low sensitivity to /spa/-/søpa/ unexpected. Also, as front-to-back clusters are expected to show tighter timing than back-to-front clusters (e.g., Chitoran et al. 2002), lower sensitivity in labial-/l/ than velar-/l/ is a bit surprising, though the perceptual recoverability argument may not be directly relevant to liquid-final clusters. Note, however, that the perceptual patterns are expected to reflect both the timing relation in the stimuli and the Georgian listeners' knowledge about the typical timing patterns in their language. Georgian listeners' knowledge about the timing would likely be gradient and have a wide range of timing patterns, as mentioned in Section 1.3.1. And this gradient variability seems to be responsible for the perceptual patterns we report in this study. In addition, we do not know whether the gestural organizations in our stimuli were consistent with these previous findings, as we do not have articulatorily measured timing data of the stimuli. It needs to be confirmed in a future study whether articulatory timing in a specific stimulus is directly reflected in the listeners' perception.

It is also interesting to note that the voicing of the component consonants, which is determinant of the appearance of the transitional vocoid in Georgian, does not seem to strongly influence the Georgian listeners' sensitivity to the CCa-CøCá

contrast. This provides further evidence that Georgian speakers' phonetic knowledge about word-initial CC clusters would involve the timing relations rather than the transitional vocoid. In addition, French listeners' discrimination of the same CCa-CøCá contrast does not show the influence of the CC composition, further indicating that Georgian listeners' difficulty may not be entirely due to the stimuli. If, for instance, /spa/ and /søpá/ had sounded more similar to each other than /ska/ and /søká/, French listeners may have also showed less accurate discrimination of the former than the latter. We did not find evidence for this (as reported in Section 3). The acoustic measurements of the first vowels in CøCá stimuli (provided in the Supplementary Material) also do not reveal any patterns that would predict Georgian listeners' behaviors.

4.3 Concluding remarks

We examined the discrimination of French CCa-CVCá pairs by Georgian listeners. The low discrimination accuracy of French pairs involving CøCá tokens by Georgian listeners suggests that they not only assimilated the nonnative vowel /ø/ to native vowel /u/, but also perceived illusory clusters when hearing French CøCá sequences. These findings demonstrate that the typical timing patterns in the listeners' native language can influence the process of perceptual modification. Listeners have knowledge about how onset clusters are temporally implemented in their native language (i.e., the typical timing between articulatory gestures), and we claim that the temporal organization constitutes language-specific knowledge that influences what may or may not operate as the target of perceptual assimilation in nonnative speech perception.

Despite the universal tendencies in the temporal organization of word-onset CC clusters (Pouplier et al. 2022), aspects of inter-consonantal timing within onset clusters are language-specific. For instance, sibilant-initial clusters have short lags across languages including both French and Georgian (Pouplier et al. 2022). However, while French speakers are familiar only with the clusters with a quick transition, Georgian speakers are familiar with more variable (from shorter to longer) lag values. That is, the range of inter-gestural timing that can be associated with a word-onset CC cluster is language-specific, and it should be learned during the process of language acquisition. Speakers acquire, as part of their phonetic grammar, the overall timing range within which they accommodate specific consonantal gestures in word-onset CC clusters. And our findings are consistent with the interpretation that this phonetic knowledge on temporal organization can influence perceptual modification patterns.

Acknowledgments: This manuscript has benefitted greatly from suggestions from Alexei Kochetov, Kevin Roon, Karthik Durvasula, and an anonymous reviewer. The authors also thank Marianne Pouplier, Pierre Hallé, Phil Hoole, and Tom Lentz for the discussions related to this study; Ramaz Kurdadze, Natia Botkveli, Maka Tetradze, and Ivane Lezhava (in Tbilisi), as well as Alice Ding (in Paris), for their help with data collection; and our Georgian and French participants for making this study possible.

Research funding: This work was supported by ANR-DFG grant ANR-14-FRAL-0004 for the project *Paths to Phonological Complexity: Onset clusters in speech production, perception and disorders*, and by the program *Investissements d'Avenir* (ANR-10-LABX-0083-LabEx EFL) contributing to the IdEx Université de Paris – ANR-18-IDEX-0001. The open access fee is funded by Seoul National University.

Author contributions: The authors confirm contribution to the paper as follows. Harim Kwon: study design, data collection, analysis, interpretation, writing, revising, editing, funding acquisition. Ioana Chitoran: conceptualization, study design, data collection, revising, editing, funding acquisition.

Conflict of interest statement: The authors have no conflicts of interest to declare.

Ethics statement: Our experiments were performed in accordance with ethical research procedures at the Laboratoire Clillac-ARP (EA 3967) Université Paris Cité, requiring participants' informed consent of their participation in the study, in compliance with the Helsinki Declaration. All participants were thus informed beforehand, in writing, of the nature and purpose of the study, and of the anonymity of their data.

References

- Adda-Decker, Martine, Cédric Gendrot & Noël Nguyen. 2008. Contributions du traitement automatique de la parole à l'étude des voyelles orales du français. *Traitement Automatique des Langues* 49(3). 13–46.
- Aronson, Howard I. 1982. *Georgian: A reading grammar*. Columbus, OH: Slavica Publishers.
- Aronson, Howard I. 1991. Modern Georgian. In John A. C. Greppin (ed.), *The Indigenous languages of the Caucasus*, vol. 1. Alice C. Harris (Ed.), *The Kartvelian languages*, 221–234. Delmar, NY: Caravan Books.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104(3). 591–630.
- Berent, Iris, Tracy Lennertz, Paul Smolensky & Vered Vaknin-Nusbaum. 2009. Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26(1). 75–108.
- Best, Catherine T. 1995. A direct realist perspective on cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171–204. Timonium, MD: York Press.

- Best, Catherine T. & Pierre A. Hallé. 2010. Perception of initial obstruent voicing is influenced by gestural organization. *Journal of Phonetics* 38(1). 109–126.
- Boersma, Paul & David Weenink. 2021. *Praat: Doing phonetics by computer* [Computer program]. Version 6.1.50. Available at: <http://www.praat.org/>.
- Bombien, Lasse & Philip Hoole. 2013. Articulatory overlap as a function of voicing in French and German consonant clusters. *Journal of the Acoustical Society of America* 134(1). 539–550.
- Borise, Lena. 2023. Disentangling word stress and phrasal prosody: A view from Georgian. *Phonological Data and Analysis* 5. 1–37.
- Borise, Lena & Xavier Zientarski. 2018. Word stress and phrase accent in Georgian. *Proceedings of Tonal Aspects of Languages* 6. 207–211.
- Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49(3–4). 155–180.
- Butskhrikidze, Marika. 2002. *The consonant phonotactics of Georgian*. Netherlands Graduate School of Linguistics, LOT 63.
- Chitoran, Ioana, Louis Goldstein & Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in laboratory phonology* 7, 419–448. Berlin: Mouton de Gruyter.
- Crouch, Caroline. 2022. *Postcards from the syllable edge: Sonority and articulatory timing in complex onsets in Georgian*. UC Santa Barbara PhD dissertation.
- Crouch, Caroline, Argyro Katsika & Ioana Chitoran. 2023a. Sonority sequencing and its relationship to articulatory timing in Georgian. *Journal of the International Phonetic Association*. 1–24. <https://doi.org/10.1017/S0025100323000026>.
- Crouch, Caroline, Ionan Chitoran, Louis Goldstein & Argyro Katsika. 2023b. Intrusive vocoids and syllable structure in Georgian. In Radek Skarnitzl & Volín Jan (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences, 2000–2004*. Guarant International.
- Davidson, Lisa. 2005. Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics* 19(6–7). 619–633.
- Davidson, Lisa. 2011. Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 37(1). 270–282.
- Dell, François. 1995. Consonant clusters and phonological syllables in French. *Lingua* 95. 5–26.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6). 1568–1578.
- Feldman, Naomi H., Thomas L. Griffiths & James L. Morgan. 2009. The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review* 116(4). 752–782.
- Fougeron, Cécile, Cédric Gendrot & Audrey Bürki. 2007. On the acoustic characteristics of French schwa. In *Proceedings of the 16th International Congress of Phonetic Sciences ICPhS XVI*, 6–10 August 2007, Saarbrücken, Germany.
- Fowler, Carol A. 1996. Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99(3). 1730–1741.
- Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural language & linguistic theory* 20(2). 269–337.
- Gafos, Adamantios & Louis Goldstein. 2012. Articulatory representation and organization. In Abigail C. Cohn, Cécile Fougeron & Marie Huffman (eds.), *The Oxford handbook of laboratory phonology*, 220–231. Oxford: Oxford University Press.

- Hall, Kathleen C. & Elizabeth V. Hume. 2013. Perceptual confusability of French vowels. *Proceedings of Meetings on Acoustics* 19. 060113.
- Hautus, Michael J. 1995. Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers* 27. 46–51.
- Hewitt, George B. 1995. *Georgian: A structural reference grammar*. Amsterdam: John Benjamins Publishing.
- Hoole, Phil, Lasse Bombien, Barbara Kühnert & Christine Mooshammer. 2009. Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In Gunnar Fant, Hiroya Fujisaki & Jiaxuen Shen (eds.), *Frontiers in phonetics and speech science*, 275–287. Hong Kong: The Commercial Press.
- Jun, Sun-Ah & Cécile Fougeron. 2000. A phonological model of French intonation. In Antonis Botinis (ed.), *Intonation*, 209–242. Dordrecht: Springer.
- Jun, Sun-Ah, Chad Vicens & Ingvar Lofstedt. 2007. Intonational phonology of Georgian. *UCLA Working Papers in Phonetics* 106. 41–57.
- Kang, Yoonjung. 2011. Loanword phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *Companion to phonology*, 2258–2281. Boston: Wiley-Blackwell.
- Kochetov, Alexei, Marianne Pouplier & Minjung Son. 2007. Cross-language differences in overlap and assimilation patterns in Korean and Russian. In *Proceedings of the XVI International Congress of Phonetic Sciences*, 1361–1364.
- Kühnert, Barbara, Phil Hoole & Christine Mooshammer. 2006. Gestural overlap and C-center in selected French consonant clusters. In *7th International Seminar on Speech Production (ISSP)*, 327–334.
- Kuhl, Patricia K. 1993. Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In Bénédicte de Boysson-Bardies, Scania de Schonen, Peter W. Jusczyk, Peter McNeilage & John Morton (eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, 259–274. Dordrecht: Springer.
- Lenth, Russell V. 2020. emmeans: Estimated marginal means, aka least-squares means, R package version 1.1. Available at: <https://CRAN.R-project.org/package=emmeans>.
- Macmillan, Neil A. & Douglas C. Creelman. 2005. *Signal detection theory: A user's guide*, 2nd edn. Mahwah NJ: Lawrence Erlbaum Associates Publishers.
- Malécot, André & Gérard Chollet. 1977. The acoustic status of the mute-e in French. *Phonetica* 34(1). 19–30.
- Meunier, Christine & Robert Espesser. 2011. Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics* 39. 271–278.
- Peirce, Jonathan W., Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas K. Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1). 195–203.
- Pouplier, Marianne, Tomas O. Lentz, Ioana Chitoran & Philip Hoole. 2020. The imitation of coarticulatory timing patterns in consonant clusters for phonotactically familiar and unfamiliar sequences. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11. 1.
- Pouplier, Marianne, Manfred Pastätter, Philip Hoole, Stefania Marin, Ioana Chitoran, Tomas O. Lentz & Alexei Kochetov. 2022. Language and cluster-specific effects in the timing of onset consonant sequences in seven languages. *Journal of Phonetics* 93. 101153.
- R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Robins, Robert H. & Natalie Waterson. 1952. Notes on the phonetics of the Georgian word. *Bulletin of the School of Oriental and African Languages* 14(1). 55–72.
- Shosted, Ryan K. & Vakhtang Chikovani. 2006. Standard Georgian. *Journal of the International Phonetic Association* 36(2). 255–264.

- Simpson, Adrian J. & Mike J. Fitter. 1973. What is the best index of detectability? *Psychological Bulletin* 80(6). 481–488.
- Stevens, Kenneth N. & Morris Halle. 1967. Remarks on analysis by synthesis and distinctive features. In Weiant Waltham-Dun (ed.), *Models for the perception of speech and visual form*, 88–102. Cambridge MA: MIT Press.
- Strange, Winifred. 2011. Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics* 39(4). 456–466.
- Tschenkeli, Kita. 1958. *Einführung in die Georgische Sprache*. Zurich: Amirani Verlag.
- Tyler, Michael D., Catherine T. Best, Alice Faber & Andrea G. Levitt. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71(1). 4–21.
- Vaissière, Jacqueline. 1983. Language-independent prosodic features. Prosody: Models and measurements. In Anne Cutler & Robert D. Ladd (eds.), *Prosody: Models and measurements*, 53–65. Berlin, Heidelberg: Springer.
- Verde, Michael F., Neil A. Macmillan & Caren M. Rotello. 2006. Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of A_z and A . *Perception & Psychophysics* 68. 643–654.
- Vicenič, Chad & Sun-Ah Jun. 2014. An autosegmental-metrical analysis of Georgian intonation. In Sun-Ah Jun (ed.), *Prosodic typology II: The phonology of intonation and phrasing*, 154–186. Oxford: Oxford University Press.
- Vogt, Hans. 1958. Structure phonémique du géorgien. *Norsk Tidsskrift for Sprogvidenskap* 18. 5–90.
- Zsiga, Elizabeth C. 2003. Articulatory timing in a second language. Evidence from Russian and English. *Studies in Second Language Acquisition* 25(3). 399–432.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/phon-2023-2005>).