Laboratory Phonology
Journal of the Association for Laboratory Phonology

Pouplier, M., et al. 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): 1, pp. 1–41. DOI: https://doi.org/10.5334/labphon.195

JOURNAL ARTICLE

The imitation of coarticulatory timing patterns in consonant clusters for phonotactically familiar and unfamiliar sequences

Marianne Pouplier[1], Tomas O. Lentz[1,2], Ioana Chitoran[3] and Philip Hoole[1]

[1] Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität, Munich, DE

[2] Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, NL

[3] Clillac-ARP, Université Paris Diderot, Paris, FR

Corresponding author: Marianne Pouplier (pouplier@phonetik.uni-muenchen.de)

This paper investigates to what extent speakers adapt to unfamiliar consonant cluster timing patterns. We exploit naturally occurring consonant overlap differences between German and Georgian speakers' productions to probe the constraints that language-specific patterns put on the flexibility of cluster articulation. We recorded articulography data from Georgian and German speakers imitating CCV clusters as produced by a German and Georgian audio model, respectively. The German participants adapted their relative overlap towards the Georgian audio model to various degrees depending on whether the cluster was phonotactically familiar to them or not. A higher degree of adaptation was observed for clusters phonotactically illegal in German. Phonotactically legal clusters showed only an intermediate degree of articulatory adaptation, even though acoustically these clusters showed a rather strong move towards the Georgian audio model in terms of the aerodynamics of the interconsonantal transition period. Georgian speakers on the other hand failed to adapt to the German audio model articulatorily and acoustically, possibly because the German cluster inventory is a subset of the Georgian inventory. This means that Georgian speakers can draw on native speaker knowledge for all clusters, which is a factor known to constrain imitation. Also language-specific cue weighting effects may partly condition the results.

Keywords: coarticulation; imitation; clusters; L2; articulatory timing; German; Georgian

1. Introduction

Speakers are able to modulate the phonetic detail of their own speech in order to resemble the speech that they just heard (Goldinger, 1998). Numerous studies have demonstrated this adaptive ability for a range of phonetic parameters such as vowel formants, duration, VOT, as well as tonal alignment (Delvaux & Soquet, 2007; D'Imperio, Cavone, & Petrone, 2014; Nielsen, 2011; Sancier & Fowler, 1997; Shockley, Sabadini, & Fowler, 2004); at the same time adaptation is constrained by social (Babel, 2012), but also by lexical and phonological factors. For instance, a study by Nielsen (2008) indicated that the scope of adaptation interacts with and is limited by the inventory of contrasts specific to the language involved: Voice onset time in an imitation experiment was only free to vary in the direction not relevant for phonological contrast. Relatedly, Whalen, Best, and Irwin (1997) found that speakers could imitate an inappropriate ([un]aspirated stop) allophone only in non-words, not in words. Therefore, speakers' imitation of each other seems to be mediated by their knowledge of their native language and the lexicon. Grammar has a stabilizing function in that phonological contrast limits the direction and amount of

adaptation. The flexible adaptation in speech production is at the same time essential for the dynamic evolution of speech, since it is this basic skill that also enables learning of new sounds, phonotactic sequences, and other languages and it is also an essential ingredient in the evolution of sound change (e.g., Harrington, Kleber, Reubold, Schiel, & Stevens, 2018; Zellou, Dahan, & Embick, 2017).

In our current study, we focus on the learnability and flexibility of coarticulatory timing patterns (patterns of temporal overlap) between successive consonants. While consonant timing is usually not considered to be a locus of category contrast, it is well known that languages generally have different constraints on how successive articulatory gestures of consonants or consonants and vowels overlap in time (Bombien & Hoole, 2013; Hermes Mücke, & Auris, 2017). We refer to this overlap as coarticulatory timing (for an overview of coarticulatory patterns arising from the temporal overlap see e.g., Farnetani & Recasens, 2010). Coarticulatory timing is part of native speaker knowledge and can be seen as part of a language-specific grammar of coarticulation. While quite a number of publications are concerned with phonotactic learning in perception and production (among many others, Davidson, 2006; Goldrick & Larson, 2008; Redford, 2008; Seidl, Onishi, & Cristia, 2013), less is known about how flexibly adult speakers can adapt the phonetic detail of consonant sequences to unfamiliar coarticulatory timing patterns, and it is this question that our study addresses.

The sequencing of consonants (and of consonants and vowels) is language-specific at two levels. First, the phonology of a language restricts which phonemes can follow each other. Such phonotactic restrictions are psychologically real as language users show knowledge of them (Moreton, 2002; Polivanov, 1931; Scholes, 1966) and use them in speech processing (Lentz & Kager, 2015; Massaro & Cohen, 1983; McQueen, 1998; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997; Weber & Cutler, 2006); they also shape L2 learning (e.g., Hallé & Best, 2007; Hallé, Segui, Frauenfelder, & Meunier, 1998). Second, the same linear sequence of segments differs between languages in how it is produced, among others, in terms of its coarticulatory timing pattern (Pouplier, 2012; Yanagawa, 2006; Zsiga, 2003). That languages contrast in the degree of coarticulation has been recognized at least since the foundational work of Öhman (1966), and the relevance of this fact for second language learning on the one hand and sound change on the other has been discussed for instance in Flege (1988) or Carignan (2018). Language-specific effects in coarticulation have been found in both vowel-to-vowel (Beddor, Harnsberger, & Lindemann, 2002; Ma, Perrier, & Dang, 2015; Manuel, 1990; Mok, 2010, 2012; Smith, 1995), consonant-vowel (Boyce, 1990; Clumeck, 1976; Lubker & Gay, 1982; Solé, 1995; Steinlen, 2005), as well as consonant-consonant coarticulation (Bombien & Hoole, 2013; Hermes et al., 2017; Kochetov, Pouplier, & Son, 2007; Yanagawa, 2006; Zsiga, 2003). Because the temporal orchestration of successive segments is language specific, coarticulatory timing has to be learned and is part and parcel of native speaker knowledge and a language's phonotactics (Davidson, 2006; Gafos, 2002).

The particular coarticulatory timing differences that are at the focus of this paper relate to the amount of temporal overlap between successive consonants forming an onset cluster, a known locus of cross-linguistic differences: For instance, the two consonants of an onset cluster overlap more in German compared to French (Bombien & Hoole, 2013). In languages such as Spanish (Hall, 2006), Norwegian (Endresen, 1991), Russian (Zsiga, 2003), or Georgian (Chitoran, 1998; Chitoran, Goldstein, & Byrd, 2002), consonant sequences may be produced with a so-called open transition. Such an open transition arises if the constriction for the second consonant is still being formed while the constriction for the first consonant has already been released (Catford, 1985). An open transition can be defined as a period of sound radiation from an open vocal tract between two constrictions (see e.g., **Figure 10b**).

This may under certain circumstances, depending on glottal aperture and aerodynamic conditions, lead to the emergence of an excrescent vocoid. It is generally understood that the acoustics of the transition are purely contextually conditioned but separating transitional vocoids from epenthetic vowels with a vocalic constriction target is notoriously difficult (e.g., Davidson, 2005; Ridouane, 2008). The work we report here does not crucially depend on such a distinction, and we return to this point in the Discussion.

In our current study, we exploit crosslinguistic differences in the temporal orchestration of consonant sequences in order to further our understanding of how flexible language-specific coarticulatory timing patterns are in an imitation task for both known and unknown consonant clusters. While previous studies have investigated how L1 speakers of a language prohibiting complex onsets produce these unfamiliar structures in L2 (e.g., Wilson, Davidson, & Martin, 2014), or how native knowledge of cluster production extends to phonotactically unknown consonant clusters in L2 (e.g., Davidson, 2005; Davidson, Martin, & Wilson, 2015), there is comparatively little work on the plasticity of coarticulatory timing patterns for segmentally identical sequences. Our current study compares the possible adaptation of such timing patterns between phonotactically known and unknown sequences. The study aims to contribute to our knowledge of how coarticulatory timing patterns are acquired and how they may flexibly adapt to unfamiliar patterns through imitation. This in turn will be relevant for our understanding of the nature of possible phonetic biases in coarticulatory timing patterns, and possible sources of stability and change in consonant clusters.

There is very little previous work on the imitation of coarticulatory timing patterns specifically. Much relevant previous work on learning of timing patterns has been done in the context of L2 research. Thereby one extensively studied area is VOT (for an overview see Zampini, 2008), which arises in part from different oral-laryngeal coarticulatory timing patterns in interaction with aerodynamics (Hirose & Gay, 1972). The case of VOT is somewhat different from the coarticulatory timing differences in consonant cluster production under investigation here, in the sense that VOT is one of the key cues to segmental contrast, whereas consonant cluster timing is not usually involved in contrast and category formation in the same way (although of course Catford, 1985, argued for such a view). Studies investigating the flexibility of L1 VOT patterns in L2 production found that speakers' L2 productions are often characterized by an intermediate VOT pattern between their native L1 and a given L2 (e.g., Flege & Hillenbrand, 1984; Flege, Munro, & MacKay, 1995; Sancier & Fowler, 1997) which suggests first that native coarticulatory timing patterns can in principle change towards non-native timing patterns, and secondly, that this change towards L2 may be partial—timing patterns of L1 may persist into L2 for many years (Flege, 1987). Relatedly, Solé (1997) specifically called attention to articulatory timing differences between Catalan and English, showing how native coordination patterns which have the status of primitives for native speakers may shape L2 pronunciation in terms of voicing, vowel nasalization, and assimilation patterns. Overall these L2 studies thus point to the fundamental flexibility of coarticulatory timing in adults while at the same time revealing limits of this flexibility.[1]

[1] There has been a surge of studies on vowel nasalization over the last decade, particularly with a focus on inter-speaker differences in the production and perception of vowel-nasal coarticulation in the context of sound change (Beddor, 2009; Beddor et al., 2002; Harrington et al., 2018; Zellou et al., 2017; Zellou & Tamminga, 2014), yet the focus of these studies is different from ours. These studies were able to show that the perception of phonetic detail is idiosyncratic and depends on experience: That is, listeners may not agree on how much coarticulation to attribute to the source that gives rise to it. Given the propensity for interlocutors to imitate each other's fine phonetic detail (Babel, McGuire, Walters, & Nicholls, 2014) sound change may be due to listeners' variable parsing of coarticulatory cues from the signal leading to variability in imitation of each other's production patterns (Beddor, Coetzee, Styler, McGowan, & Boland, 2018).

One of the few studies specifically addressing the question of coarticulatory flexibility in the context of L2 was presented by Steinlen (2005) in a comparative study of consonant-vowel coarticulation in Danish, British English, and German. Based on acoustic recordings she first confirmed differences in the degree of CV coarticulation between these languages. In particular Danish vowels showed hardly any spectral effects of consonantal context, while English and German vowels were affected by context to a similar degree. Despite the virtual absence of coarticulatory effects of consonants on vowels in Danish, Steinlen's Danish participants produced a high degree of CV coarticulation in L2 English, but only in some consonantal contexts, not in others. German speakers did not adapt their CV coarticulation in L2 English. Steinlen concluded that while L2 speakers transfer language-specific aspects of coarticulation from L1 to L2, there was also some evidence for the learnability of the L2 pattern in her study (the Danish speakers adopting some of the English coarticulatory patterns).

Davidson (2005, 2006) pioneered the use of articulatory measures to probe coarticulatory timing patterns in unfamiliar consonant cluster production. Her work revealed that production inaccuracy for unknown phonotactic sequences is at least partially rooted in insufficient temporal overlap: English native speakers articulated the component consonants of unfamiliar clusters at least some of the time temporally so far apart from each other that a vocoid emerged between the consonants. She suggested that what may look like vowel epenthesis may in fact be the acoustic result of an open transition between two consonants, arising from what she called 'gestural mistiming.' English speakers thus produced phonotactically illegal sequences with a timing pattern different from the native English pattern, since English onset cluster productions are not typically characterized by an open transition. While speakers did not retain their native timing pattern, they neither adopted the pattern they heard for the foreign CC sequences, since the CC sequences were produced without a vocoid (and in fact contrasted with CəCV in the same experiment). Davidson argued that gestural mistiming arises not just from a lack of motor skill, but rather that, for English, phonotactic constraints comprise a prohibition on close coordination patterns for phonotactically illegal sequences. That is, for English L1 speakers, a high overlap pattern is part and parcel of phonological phonotactic legality: If abstract gestural coordination is viewed as part of phonological representations in the vein of Gafos (2002), phonotactic illegality extends to a prohibition of producing illegal sequences with the native gestural coordination pattern. Our research extends this work by comparing the imitation of phonotactically illegal clusters with the imitation of known clusters with an unfamiliar coarticulatory timing pattern which may include an open transition. If known clusters are imitated within one's native language with a different timing pattern, this would provide evidence for the possibility of open transitions arising as part of faithful imitation.

In terms of the characteristics of non-native production patterns, two major factors are discussed in the literature. For one, it could be shown that the specific type of response produced for auditorily presented phonotactically illegal sequences will vary as a function of fine phonetic detail of the stimuli (Wilson & Davidson, 2013; Wilson et al., 2014). For instance, Wilson and Davidson (2013) systematically manipulated the phonetic detail of their stimuli and confirmed that increased burst intensity conditioned higher rates of epenthesis in illegal CCV sequences. Lower amplitude bursts also predicted a higher rate of C1 deletion or misidentification (see also Daland, Oh, & Davidson, 2018 for phonetic factors in loanword adaptation).

Secondly, to study speakers' performance with unfamiliar sound patterns is a well-known approach to assessing universal knowledge and constraints which presumably emerge in

situations in which learned behavior is not available without transfer (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008; Wiese, Orzechowska, Alday, & Ulbrich, 2017). The observation that not all patterns observed in L2 are explainable on the basis of transfer serves as a main argument for the role of markedness in L2 learning. For one, there may be asymmetries in transfer, such as observed among others by Zsiga (2003), a study we will review in detail below. Asymmetries in transfer may arise if a given L2 has a universally less marked pattern compared to a given L1. In this case the L2 pattern will be learned relatively easily, but not vice versa (e.g., Eckman, 2008). If it is L1 that has an unmarked structure, transfer of the unmarked pattern to L2 is likely (Cebrian, 2000). Likewise, it has been argued that even in the total absence of a particular structure in L1 (such as consonant clusters), universally unmarked structures are acquired more readily than marked structures (e.g., Berent, Steriade, Lennertz, & Vaknin, 2007; Broselow, Chen, & Wang, 1998).

Naturally occurring overlap differences between phonotactically identical CC sequences, as we employ in our study, have only very limitedly been exploited before, either in L2 or in imitation studies. Zsiga (2003) is to our knowledge one of the first studies in this area, specifically investigating the L1 to L2 transfer of coarticulatory timing patterns in interaction with markedness constraints. She analyzed acoustic recordings of stop-stop consonant sequences across word boundaries contrasting English and Russian. Zsiga specifically tested whether consonant timing in these sequences (e.g, English *sto**p** **t**arts*) would be characterized by a transfer of the L1 pattern or by the emergence of an unmarked pattern in which speakers would be expected to maximize perceptual recoverability by pronouncing each word as a separate unit ('word integrity effect,' see also Cebrian, 2000). In English, there is often substantial overlap in stop-stop sequences and C1 in C1#C2 sequences is mostly unreleased.[2] This contrasts with an obligatory release of word-final stops in Russian, indicative of a lesser degree of overlap across word boundaries (Kochetov et al., 2007; Kochetov & So, 2007; Zsiga, 2000).

Zsiga found that Russians continued to release the first consonant even in L2 English, keeping their native pattern in L2. English speakers, on the other hand, adopted the Russian pattern, but did not differentiate between cluster types as Russian L1 does. Zsiga interpreted the across-the-board low overlap pattern for English speakers within a perceptual optimization approach: Speakers optimize cue recoverability in L2 on a word-by-word basis (Weinberger, 1994). According to Zsiga, English natively sacrificies perceptual recoverability of final consonants in stop-stop sequences (though see Gow, 2002) in favor of cueing prosodic phrasing, whereas Russian ranks the requirement to release C1 higher. Since recoverability principles seem to dominate non-native language production, for Russians the native and unmarked non-native pattern coincide, but English speakers will adopt a pattern similar to L1 Russian in an instance of the emergence of the unmarked.

English speakers' behavior could be interpreted as emergence of the unmarked (rather than just adoption of the Russian pattern) because English speakers failed to implement cluster-specific timing differences which are, according to Zsiga, characteristic of Russian but not English. Zsiga thus found effects of both transfer of an L1 pattern and the emergence of the unmarked. Yet Yanagawa (2006) who systematically compared consonant-consonant timing for consonant sequences in various positions for four languages could not replicate Zsiga's results: In an articulography study, she compared L1 Cantonese, Japanese, and

[2] Davidson (2011) reports that 25% of stops followed by a consonant are released in spontaneous speech; see also Byrd and Tan (1996).

German speakers but found that language-specific differences in CC timing failed to predict L2 English production. All speakers adopted a pattern quite close to that produced by control English native speakers. She thus could not replicate for her languages the results Zsiga (2003) had obtained for Russian and English. Since Zsiga (2003) assessed articulatory overlap indirectly from acoustic measures, possibly more subtle effects may have been missed, especially for stop-stop sequences. For our current imitation study we recorded articulatory data for two speaker groups with different coarticulatory timing patterns, and thus will be able to observe fine-grained effects of articulatory adaptation.

In our current study, we capitalize on the coarticulatory timing differences between Georgian and German as they have been reported in the literature. Georgian has been described as having a low degree of overlap between consonants (Chitoran et al., 2002), while German has been described as having a high degree of overlap in consonant clusters (Bombien & Hoole, 2013; Pouplier, 2012). The goal of our current study is thereby to shed light on the adaptive flexibility of coarticulatory timing patterns for known and unknown clusters, and on whether this flexibility depends on the specific coarticulatory timing pattern of a given language. This allows us, in turn, to gain further insights into possible phonetic biases in coarticulatory timing patterns. The particular focus is thereby on the German speakers for whom we contrast clusters which exist phonotactically in both languages—but with different coarticulatory timing patterns—with those that exist in Georgian only. If German speakers imitate, within a German speaking context, a Georgian-like low overlap pattern for known clusters, this would provide evidence that low overlap patterns may arise as part of a controlled imitation of an unfamiliar pattern within one's native language. If, on the other hand, imitation or move to a generic low-overlap pattern occurs for unknown clusters only, this would provide evidence for native coarticulatory timing patterns being deeply entrenched for phonotactically legal clusters vis-à-vis the force of phonetic biases in unknown cluster production.

For the Georgian speakers of our experiment, on the other hand, all clusters are known. Since it has been shown that access to native language knowledge limits imitation (Nye & Fowler, 2003), we could expect less imitation. The same prediction of a lesser degree of imitation by the Georgians could be gleaned from Zsiga's L2 study on the grounds that Georgian has been reported to have a low overlap pattern (which Zsiga argues to be unmarked). Since the timing patterns of Georgian have not been extensively studied, we formulate this possible prediction with care. On the other hand, our Georgian participants were recruited in Munich and have some L2 knowledge of German which would give reason to expect a faithful imitation of German.

Before presenting our Methods, we will now briefly contrast the main characteristics of Georgian and German relevant for our study. Georgian is a Kartvelian language of the Caucasus which has an extensive cluster inventory (Butskhrikidze, 2002; Shosted & Chikovani, 2006). For stop-stop sequences, an obligatory release has been reported (Chitoran, 1998), consistent with the observation of a relatively low C-C overlap pattern. German has a more restricted cluster inventory (Wiese, 2000). In contrast to Georgian, which has no central vowels, German allows for vowel reduction in unstressed weak syllables and has the central vowels /ə, ɐ/. Thereby especially Southern German, the region from which we recruited our speakers, is characterized by schwa elision in the pretonic prefixes */bə-, gə-/* unless the stem is stop-initial, in which case the entire prefix is dropped (Rowley, 1990; Wiesinger, 1989).

The stop phonation contrast in Georgian is described phonologically as voiced, voiceless aspirated, and laryngealized. Prevoicing in Georgian is weak though and there have been suggestions to better cast the contrast in terms of unaspirated, aspirated, and laryngealized

(Butskhrikidze, 2002; Shosted & Chikovani, 2006). The German stop contrast is described as voiced/lenis versus voiceless aspirated/fortis. Fortis stops can be aspirated in all positions, except when following a sibilant in a cluster (Hoole, 2006; Kohler, 1995). Voiced stops show considerable variation depending on position, context, region, and speaker idiosyncracies and can in initial position variably be prevoiced (particularly in voiced contexts) or voiceless unaspirated (Jessen, 1998; Piroth & Janker, 2004). We infer that the phonation of phonologically voiced stops in German and Georgian may be relatively similar, although to our knowledge they have never been directly compared in this respect.

In sum, given the known global differences in coarticulatory timing between German and Georgian, and given that listeners are very sensitive to coarticulatory variability, we explore to what extent native coarticulatory timing patterns are flexibly adapted in a task designed to elicit the imitation of non-native timing patterns, and whether there is evidence for low overlap patterns arising independently of functional pressures. While previous research on English leads us to expect that coarticulatory timing patterns will be stable for known clusters but move to a generic low overlap pattern for unknown clusters, Zsiga's results obtained for Russian in the context of L2 research point to the possibility that also in an imitation task the degree of adaptation might not be equal across languages. While we recorded German as well as Georgian speakers imitating each other's CC coarticulatory timing patterns, the emphasis of our study is on the German speakers since it is only for this speaker group that we can contrast known and unknown sequences and that we have a relatively clear picture of native coarticulatory timing patterns. In addition, our Georgian participants already had some exposure to German.

Methodologically, we investigate CC coarticulatory timing patterns by relating specific articulatory 'landmarks' to each other, a method which has been used in previous research to quantify overlap patterns in consonant sequences. These will be described in detail below. Although these measures can be used as a convenient tool to differentiate coarticulatory timing patterns, it is possible that participants zoom in on other aspects of non-native productions when imitating them. To determine if participants imitate the non-native coarticulatory timing patterns they hear in any way, we additionally employ a greedy machine learning approach to separate Georgian and German speakers' timing patterns on whichever aspect they differ. Such a global separation then allows us to see whether there is imitation in other dimensions than those quantified by our measures. In addition, we also, for a subset of our stimuli, ask to which extent the acoustic phonation profile of the transition between consonants is imitated.

2. Methods

The recordings of two audio models served as stimuli for the main experiment. For the main experiment, we had two groups of speakers, one group of German native speakers and one group of Georgian native speakers. The technical recording setup was identical for all parts.

2.1. Audio model recordings and stimulus preparation

One female native speaker of German, aged 25, and one female native speaker of Georgian, aged 26, served as audio models. The German audio model was a member of the Institute of Phonetics in Munich but naive as to the questions of the experiment. The Georgian audio model was selected on the basis of her being the first Georgian native speaker to respond to our search for participants and realizing a coarticulatory timing pattern

different from the German speaker in the direction expected given the existing literature. She was a fluent L2 speaker of German.[3]

The audio models were asked to read, in their native language, CCa non-words in a neutral carrier sentence displayed on a computer screen. All CC combinations were phonotactically legal in the native language of the respective audio model. The German phrase was /ɪç hø:ʁə CCa: bɛsɐ ap/, *Ich höre CCa besser ab*, meaning 'I'd better check CCa by listening,' and /sitʼqʼva CCa pʰutʃʼia/, for Georgian, meaning 'The word CCa is useless.'

For all participants, audio models and others, we recorded articulography data (EMA) with synchronous audio recording (details below). From the audio models' data the CCV syllables were excised from the carrier sentence. For each cluster and each audio model, the last of two (Georgian) or three (German) repetitions was selected, unless the last repetition was erroneous or noisy, in which case the penultimate was used. This token selection was performed in this way in order to maximize familiarity with the stimulus material and the recording situation. The speakers were also familiarized with the recording material prior to the beginning of the recording session. For the German audio model recordings, the vowel of all tokens was reduced by 30 ms to remove coarticulation with the following labial of the carrier phrase which induced a /CCVb/ percept in the excised syllables. For the Georgian stimuli this was not necessary since no labial coda percept was induced by the carrier phrase, as confirmed by informal pilot testing. All stop-initial clusters were cut at the beginning of the burst. All recordings were intensity equalized using Praat (Boersma & Weenink, 2018).

2.2. Main experiment: Participants

For the main experiment, eight native speakers of German without knowledge of Georgian and living in the Munich area participated in the German-native version of our experiment. An additional two German participants were tested but had to be excluded, one for technical recording reasons and one for failure to complete the task as intended. The German participants had spent most of their lives in Germany and reported L2 knowledge in various Germanic and Western-Romance languages (one of them also spoke L2 Japanese). The two speaker groups had the same median age of 26 with a range of 22–30 for the German and 20–30 for the Georgian group. For the Georgians, data from 10 native speakers are available for analysis; data from an eleventh participant had to be excluded for technical reasons. All participants resided in Munich at the time of the recording and they all had some level of L2 knowledge of German. They had lived in Germany between 4 months and 7 years at the time of recording with the average being 3.3 years (*SD* 2.4; median 3.5). The level of L2 knowledge of German was not controlled for due to the constraints on finding Georgian native speakers in Munich. Four of the participants had studied German in school, one had studied German at university level, one participant had studied German for 5 years, another for 12 years. Two others had been studying German for only 4 months; one of them had virtually no knowledge of German. Some participants were thus very fluent in German; others had only very basic knowledge. They all reported speaking Russian (two were Georgian Russian bilinguals); most of them also had L2 knowledge of English. All consent forms were translated into Georgian.

[3] We use the term *audio model* since these speakers' recordings served as stimuli for the imitation task. We thereby make no claims about where these speakers fall within the general statistical distributional properties of the overall population of German or Georgian native speakers. Crucially, what is primarily important for our experiment is that these audio model speakers realize a coarticulatory timing pattern different from each other and their timing patterns are on a macroscopic level consistent with the existing literature on coarticulatory consonant sequence timing in these languages.

2.3. Main experiment: Experimental protocol

The participants' task was to imitate as part of a constant carrier phrase the audio stimulus they heard at the beginning of each trial. The carrier phrase was in a given participant's native language. Each audio stimulus was presented twice with 500 ms silence between the two repetitions. This was done in an effort to minimize the danger of categorical misperception of the stimuli, which were presented in isolation.

The experiment consisted of a baseline *native* condition and a *non-native* test condition which were presented in two blocks. The native condition was always recorded first and consisted of the clusters phonotactically legal in the respective native language of the participants, with the audio stimuli being from the audio model sharing the given participant's native language. Thus German participants heard the German audio model, Georgian participants heard the Georgian audio model. In the non-native condition, the audio stimuli of the 'other' audio model were used, i.e., German participants were presented with the audio recordings of the Georgian audio model and Georgians listened to the German audio model's productions.

Note that for the Georgians, all clusters were phonotactically legal in both conditions, only the audio model differed between the native and non-native conditions. This is because the German #CCV cluster inventory is a subset of the Georgian inventory. For the German speakers, the non-native condition comprised all the clusters of the native condition (heard with the coarticulatory timing pattern of the Georgian audio model) plus the additional clusters which exist in Georgian only (see below). For the native conditions, three repetitions were recorded per cluster in differently randomized blocks; for the non-native conditions, there were eight repetitions. Experiment time, i.e., number of repetitions, was divided unequally between the two conditions for the following reason: Whereas the native condition merely served as a baseline, for the non-native condition, we were interested in the possibility of practice or learning effects over the course of the experiment and thus biased the repetition number in favor of this condition. Since no such effects were found for our measures, we do not report on this aspect any further.

Overall we expected that the Georgian audio model's productions should show a lower degree of C-C overlap compared to the German audio model's productions for the lag measures employed (see below). It turned out that the German and Georgian audio models' productions, while generally distinct, overlapped due to three clusters: /sp, ʃm, dg/, for which the Georgian audio model produced a higher degree of temporal overlap compared to the German audio model on at least one of our measures (we return to this observation in the Discussion). This was evaluated on a by-cluster basis for /sp/ and /ʃm/; for /dg/, which does not exist in German, the production was compared to the mean of the German audio model's productions. We excluded these three clusters altogether.

A total of seven clusters that occur in both German and Georgian and an additional fourteen that occur in Georgian but not in German were included in the analysis. Clusters existing in both languages were /ʃp, sm, bl, gl, gn, kl, kn/; clusters existing in Georgian only were /bg, bn, gb, gd, lb, lm, md, mg, ml, mn, mt, nb, ng, tm/.

The targeted data set thus contains 7 clusters × 3 repetitions × 8 speakers = 168 tokens for the German native condition. The German non-native condition comprises 21 clusters × 8 repetitions × 8 speakers = 1344 tokens. For the Georgian native condition, the data set contains 21 clusters × 3 repetitions × 10 speakers = 630 tokens; for the non-native condition 7 clusters × 8 repetitions × 10 speakers = 560 tokens, rendering a targeted grand total of 2702 tokens across both languages and conditions.

Participants were instructed that they would hear syllables from a model speaker and their task was to repeat these syllables after the audio model, embedded in a sentence,

to the best of their abilities. All participants were advised that the syllables of the first part of the experiment would sound quite familiar to them, whereas in the second part the syllables were from a different language and might sound unfamiliar. In any case they should still repeat back what they heard from the audio model as closely as possible. They were not explicitly informed about the specific languages the syllables were from in order to keep their attention focused on the phonetic detail of the stimuli; participants from both groups reported informally after the experiment that they recognized the stimuli of the first part as being from their respective native language.

The carrier phrase was displayed on the screen for each trial with a blank for the target syllable. The carrier phrase was for all conditions in the given participant's native language. The German carrier phrase was the same as for the audio model: /ɪç høːʁə CCa: bɛsɐ ap^{h}/, *Ich höre CCa besser ab*, meaning 'I'd better check CCa by listening.' The Georgian carrier phrase was for technical reasons slightly different from the one used for the audio model: /veeba CCa p^{h}utʃ'ia/, meaning 'A large CCa is useless.'

2.4. Recording procedure

We used the Carstens AG501 articulograph to record articulatory movement data for both of the audio models and the participants. We used standard calibration and position recovery procedures for every recording session. Three receiver coils were placed on the tongue: just behind the tongue tip (TT), one on the dorsum to capture velar constrictions, and one in the mid region. One coil was placed on the lower incisors to measure jaw movement and one coil each was placed on the vermilion border of the upper and lower lips. All of these coils were placed in the midsagittal plane. Reference coils to correct for head movement were placed on the nose ridge, the maxilla, and behind each ear on the mastoid process.

After position recovery the recordings were corrected for head movement and rotated to the occlusal plane. Articulatory data were sampled at 1250 Hz. As part of position recovery, the signal of each transducer was low-pass filtered with a cut-off frequency of 20 Hz, with the exception of TT which was filtered at 60 Hz and the reference coils which were filtered at 5 Hz. The velocity signals of all time series were smoothed at 24 Hz with a moving average filter. Synchronously with the articulatory data audio recordings were acquired at a sampling rate of 25.6 kHz.

2.5. Analysis

2.5.1. Auditory Analysis

Trials in which the pronunciation differed in segmental composition from the given audio stimulus (e.g., /s/ instead of /ʃ/; /nba/ instead of /bna/) were identified based on auditory evaluation, combined with visual inspection of the oscillogram, spectrogram, and the articulatory time series data. This was performed by one of the authors (T.L.). We operationally term these tokens categorical errors. No meaningful analysis of the kinematics could be performed for these tokens and they were analyzed according to error category only. This applied to 9% of the data (246 Tokens). For the remaining tokens, articulatory gestures were annotated according to the procedures detailed in Section 2.5.2.

2.5.2. Gesture annotation and measurements

For each sensor time series of interest, the articulatory constriction formation and release were identified in *Mview* (kindly provided by Mark Tiede) based on the first derivative of the position signals. For each gesture, the peak velocities associated with constriction formation and release, and the velocity minimum associated with the maximal constriction were identified. A 20% threshold of the peak velocity was used to identify the time points

of movement onset, achievement of target (plateau onset), and release (plateau offset). Thresholds were adjusted in a range between +/–5% if the automatically detected gesture was clearly incorrect. Only for two tokens one of the time points, namely the onset of the first consonant gesture, could not be reliably identified and this landmark was not used. This is only relevant for the machine learning analysis (see Section 3.2.3) as the other analyses do not use this time point. Movement onset and target achievement were determined for each consonant relative to the peak velocity of the constriction formation; the release time point was computed relative to the peak velocity of the constriction release. The time point of maximal constriction was identified as velocity minimum between the two peak velocities.

To measure labial constrictions, lip aperture was computed as the Euclidean distance between the upper lip and lower lip. Coronal constrictions were segmented based on the tangential velocity profile of the tongue tip sensor; dorsal constrictions were identified based on the first derivative of the vertical movement component of the tongue dorsum sensor.[4] Segmentation of landmarks based on the velocity profile is illustrated in **Figure 1**.

As quantification measures of gestural overlap, we employed two articulatory measures which have both routinely been employed in previous studies investigating consonant-consonant timing (among others, Chitoran et al., 2002; Marin, 2013; Pouplier, Marin,

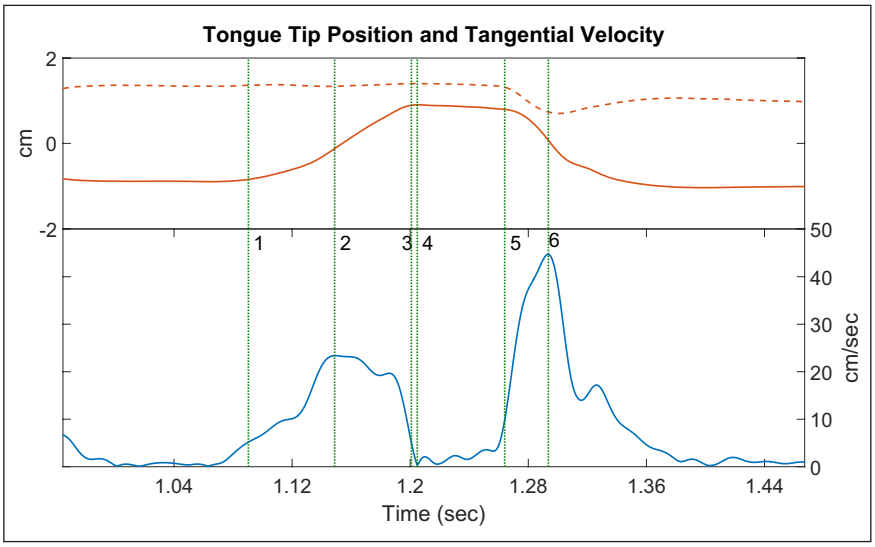

Figure 1: Illustration of kinematic segmentation of a tongue tip constriction (here, for /l/). The top panel shows horizontal (dashed line) and vertical (solid line) position of the tongue tip sensor over time. The bottom panel displays the tangential velocity profile. Numbers indicate the time points of the following landmarks: 1) movement onset of constriction formation, 2) peak velocity of constriction formation, 3) target achievement, 4) maximum constriction, 5) release, 6) peak velocity of release. See main text for details. For display purposes, the position values have been centered to zero.

[4] These measurement variables were selected since they capture the main constriction formation and release motions of the different articulatory structures (see for instance also Bombien & Hoole, 2013). In particular velars are known to show a looping motion with a fronting of the tongue during occlusion (Perrier, Payan, Zandipour, & Perkell, 2003). The tangential velocity profile is sensitive to this fronting motion and thus impacts our plateau measure in a way orthogonal to our study. For coronals, a raising and fronting is often seen as part of the constriction formation itself and as such the tangential velocity profile is an appropriate basis for measuring constriction formation, target, as well as release. An aperture function was chosen for labial constrictions since both upper and lower lip contribute to labial constriction formation and release. Importantly, the measurement variables for a given constriction were constant across all conditions and speaker groups, thus any relative differences in coarticulatory timing between conditions and speaker groups cannot be due to the choice of tract-variable specific measurement dimensions.

Hoole, & Kochetov, 2017). For one, *normalized plateau lag*, defined in (1), is the interval between release of C1 and target achievement of C2, normalized for the duration of the entire C1C2 target plateau (target achievement of C1, release of C2). Secondly, *onset lag*, as defined in (2), quantifies the time point of C2 movement onset relative to C1's target achievement onset. This measure is normalized for the plateau duration of C1 and expresses the proportion to which the plateau of C1 is produced prior to the onset of C2. Negative values mean that C2 begins its movement before C1 has reached its target, as is the case in the example given in **Figure 2**.

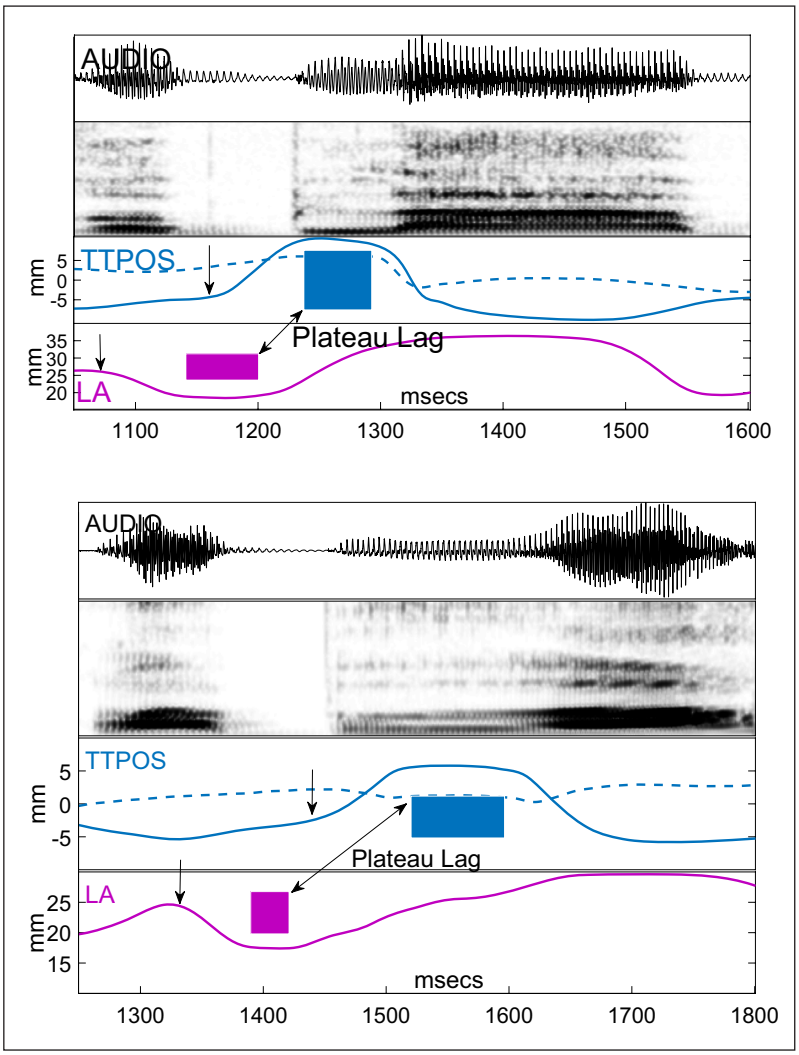

Figure 2: Illustration of coarticulatory timing measures taken from the kinematic signals. Top figure /bla/ produced by the German audio model. Bottom figure: /bla/ produced by the Georgian audio model. In each figure, the four panels show, from top to bottom: audio, spectrogram, tongue tip position (horizontal: dashed line, vertical: solid line; positions centered on zero for display purposes), lip aperture. Time is on the x-axis. The boxes indicate the plateau interval (target achievement to release) for each gesture. The arrows indicate the time points of movement onset for each consonant. Absolute plateau lag is indicated by an arrow; this interval was normalized for the interval from plateau onset of C1 to the plateau offset of C2. Onset lag was defined as the onset movement of C2 (here: tongue tip) relative to the plateau onset of C1, normalized for plateau duration of C1. All landmarks were identified based on the velocity profile; see main text for details. Note the marked difference in plateau lag and onset lag between the two productions of /bl/. The Georgian production can be characterized as having an open transition due to the pronounced lag between the acoustic release of C1 and target achievement of C2.

$$\text{Plateau lag} = \frac{\text{target achievement C2} - \text{release C1}}{\text{release C2} - \text{target achievement C1}} \quad (1)$$

$$\text{Onset lag} = \frac{\text{movement onset C2} - \text{target achievement C1}}{\text{release C1} - \text{target achievement C1}} \quad (2)$$

There were overall 26 tokens which could not be included in the kinematic analysis due to technical difficulties associated with position recovery. Additionally, speakers sometimes seemed to hesitate on the first consonant of the cluster, leading to overly long plateau values which severely impacts the onset lag measure. We therefore decided to remove outliers from the data set based on the interquartile range. Any token with an onset lag or normalized plateau lag value more than three times the interquartile range from the upper/lower quartile was considered an outlier and removed from analysis. The procedure identified 66 outliers. The final data set for the kinematic analyses thus comprises 2364 tokens. The machine learning analyses reported below only use clusters that occur in both languages, of which one token had to be discarded due to segmentation difficulties, as mentioned above (a German native production of /kl/).

2.5.3. Acoustic analysis

For acoustic analysis of the interconsonantal transition interval we first cut the CCV target words from the acoustic beginning of C1 to the end of the vowel for all files. All analyses were performed in Matlab. We centered the signal to zero and performed amplitude normalization on the CCV interval; this means effectively that all tokens were normalized to the amplitude of the vowel. Our choice of acoustic analysis parameters was guided by the concern that many of the transitions are fairly short; we therefore favored methods that are robust for short time windows over any methods that rely on pitch period detection.

We quantified the degree of voicing during the closure of C1 by taking the median of the absolute normalized amplitude of the signal during stop closure. Stronger voicing during closure is associated with higher signal amplitude values.

To identify transition intervals, we decided to use a combination of acoustically and articulatorily defined landmarks as given in (3), since demarcating a transition by visual inspection of the acoustic signal can be quite tricky.

$$\text{transition} = \text{articulatory target achievement C2} - \text{acoustic burst C1} \quad (3)$$

We used the acoustic release of C1 rather than the articulatory release, since due to tissue viscosity the articulatory release systematically precedes the acoustic release and occurs at a time point at which the acoustic signal indicates that the vocal tract is still closed, even though articulator motion out of the constriction is well on its way. The acoustic release of C1 was manually identified and defined as the beginning of the burst. Note that (3) implies that a transition interval is defined for all tokens, independently of their coarticulatory timing pattern, and that the transition interval can be negative, (close to) zero, or positive (for examples see Section 3.2.4, particularly **Figure 10**).

Phonation profile analyses were performed on the transition interval as defined in (3), but only tokens with transitions $> +10$ ms were taken into account since for shorter transitions a phonation profile analysis is not meaningful. The phonation profile of each transition $> +10$ ms was quantified on the basis of amplitude envelope modulation and normalized zero crossing rate. For voiced speech, energy tends to be concentrated

in frequencies below 4 kHz (though dominated by f0 and F1); for voiceless parts the energy is expected in the higher frequency regions. This implies higher zero crossing rates compared to voiced parts of the signal. Zero crossing rate for a given transition was normalized for the number of samples in that transition.

The amplitude envelope was calculated using Matlab's *envelope* function set to use spline interpolation over local maxima. We captured modulation of the amplitude envelope, ΔE, based on the difference between two envelope time points: the envelope amplitude at the beginning of the transition (E1) and the maximum of the envelope at least 10 ms into the transition (E2). Envelope modulation was thus defined as ΔE = E2–E1. The 10 ms threshold on the location of E2 was imposed due to some tokens—those with a very strong burst—having their envelope maximum at the beginning of the transition. Due to the threshold ΔE will be negative for such tokens. In order to avoid edge effects in the envelope amplitude values at the beginning of the transition, we averaged the envelope values of the first 2 ms to determine E1. Examples for transitions and their analysis values can be found in Section 3.2.4, particularly **Figures 12, 13**.

2.5.4. Statistical analyses

For statistical analyses, we used mixed linear models with the lme4 and lmerTest packages in R (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017; R Core Team, 2018) with a Satterthwaite approximation of the degrees of freedom.

As stated above, we recorded three repetitions for the native and eight repetitions for the non-native condition with the intention to look and control for practice effects. However, there were no meaningful repetition effects in the data for any of our analyses and we thus included the factor Repetition as a nuisance variable in the statistics but will not comment on it any further.

3. Results

3.1. Audio model cluster productions

For all audio model productions which are used in the subsequent analyses of the main experiment, **Figure 3** shows the onset lag and normalized plateau lag values. The larger filled symbols represent the audio model specific means across productions. As can be seen visually, the desired separation of two coarticulatory timing patterns was achieved.

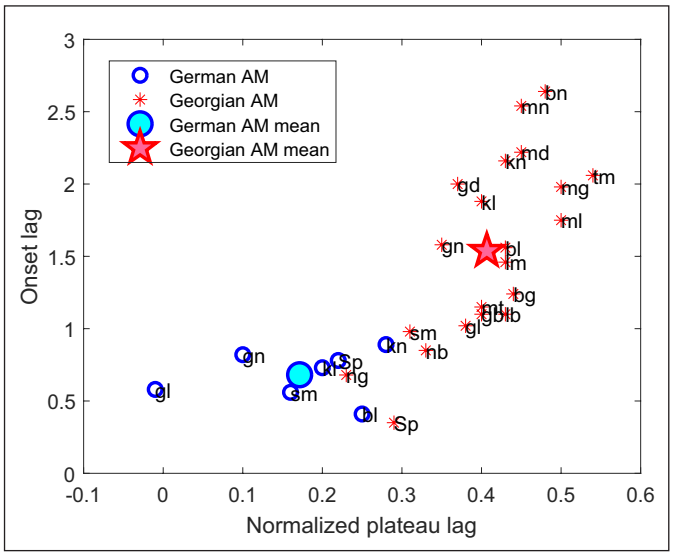

Figure 3: Scatterplot of audio model productions. Capital <S> stands for /ʃ/. AM stands for audio model.

To quantify the distance between the audio model productions statistically, we calculated the mean across all clusters separately for the German and Georgian models. The mean normalized plateau lag and onset lag were .17 and .68 for the German audio model and .41 and 1.54 for the Georgian audio model. To test whether there is a significant difference between the two audio models, we compared the set of German audio model data to the mean of the Georgian audio model and vice versa, using a t-test. For the German data, the t-tests were significant for both normalized plateau lag ($t(6) = -6.3$, $p < .001$) and onset lag ($t(6) = -13.3$, $p < .001$). For Georgian, the t-tests were likewise significant for both normalized plateau lag ($t(20) = 14.3$, $p < .001$) and onset lag ($t(20) = 6.3$, $p < .001$). Hence, the set of productions of each audio model differed significantly from the coarticulatory timing pattern of the other model, in the direction that we expected given the literature.

3.2. *Main experiment*

3.2.1. Categorical errors

There were overall 246 categorical errors, i.e., productions in which consonants were produced in incorrect order, or with a different manner or place. The majority of errors affected both consonants, followed by errors on C1. C2 was less affected by error (both: 45%, C1 only: 42%, C2 only: 13%).

We binned the categorical errors into the following categories: addition, deletion, place error, manner error, serial order, or multiple errors (e.g., manner and place or place and order).[5] **Figure 4** gives the proportional error types. Proportions were calculated relative to the token total of each native language and condition (native, non-native). Overall, both the German and Georgian native conditions had errors on 6% of tokens, the German non-native condition on 8% of tokens, the Georgian non-native condition on as many as 16% of tokens. The higher error rates for the non-native condition agrees with other studies that have found worse phoneme-identification in non-native compared to native speech (Cutler, Weber, Smits, & Cooper, 2004).

Some of the errors, particularly those occurring in the native conditions, were presumably caused by the fact that some of the auditory stimuli contained initial stops and some

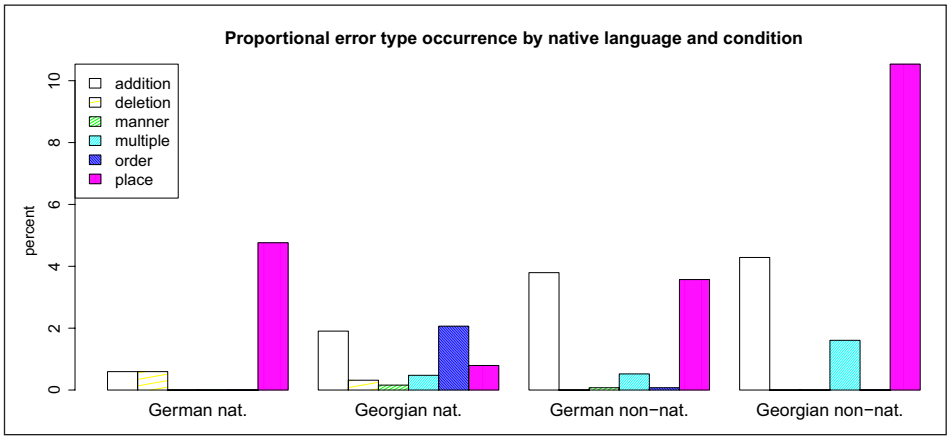

Figure 4: Error type proportions within each native language and condition.

[5] Note that the audio models differed in the duration of the transition as defined in Section 2.5.3 of the Methods (see Section 3.2.4 in Results for analysis), but the nature of the transition between the consonants produced by our speakers in the main experiment was not considered in the categorical error analysis. The categorical error analysis concerns only tokens which had to be excluded from the continuous signal analyses because the intended consonants were not produced or produced in wrong serial order, making a continuous signal analysis void. Transitions are analyzed acoustically in detail in Section 3.2.4.

stop-stop sequences, which can be expected to be perceptually vulnerable (e.g., Miller & Nicely, 1955). This problem could not be avoided, since presenting the clusters in a /V#CCV/ context might have induced perceptual resyllabification of the phonotactically illegal clusters for the Germans. Another error that occurred frequently was that sibilant-initial clusters were often pronounced with a preceding labial, such as /pʃp/ for /ʃp/. This occurred in 51 instances, thus making up 21% of all errors across conditions. This specific error makes up for 1% of tokens in the native conditions versus 2% of tokens of the non-native conditions (across native languages).

Concentrating on the errors occurring in the non-native conditions, we observed for the German speakers a predominance of place errors (41), followed by initial vowel additions (24), consonant additions (8), and a small number of various other types (order: 1, manner: 1, multiple: 4). The 24 initial vowel additions occurred exclusively in the German non-native condition for the sonorant-initial clusters /md, lb, lm, nb/ which were produced as VCCV. We note that all of these were produced by a single speaker and thus must not be attributed to a general perceptual repair strategy that emerged in the experiment overall.

For the Georgian speakers, the high number of place errors stands out in the non-native condition. Of these 59 cases, 6 are due to /s/ being pronounced as /ʃ/; all the other cases are due to /gl/ being mispronounced as /bl/. The /gl/ produced by the German audio model was repeated as /bl/ or /vl/ consistently by 7 Georgian speakers and 3 German speakers. This error also accounts for the comparatively high error rate in the Georgian non-native condition.

3.2.2. Kinematic data analysis

We now proceed to evaluate the kinematic data of the main experiment. As stated above, the categorical errors accounted for in the previous section are excluded from the analyses presented in the remainder of the paper. To assess whether speakers changed their production patterns between native and non-native conditions, we calculated the log squared Mahalanobis distance ratio of each token from the main experiment to the audio model distributions (in two dimensional space using onset lag and normalized plateau lag) as follows. We calculated the squared Mahalanobis distance of each token to the two audio model productions, giving us two Mahalanobis distance values, M1, M2, per token which represent the distance of a given token to the German (M1) and Georgian (M2) audio model productions respectively. We then take the log Mahalanobis distance ratio (MDR), as log(M1/M2), to quantify the relative proximity of the given token to the two audio model distributions.

An MDR of zero means that the token is equidistant to the two audio model distributions. A negative MDR means that a given token is closer to the German audio model than to the Georgian audio model. Our main question is whether the MDR values change between native and non-native conditions and whether German speakers further differentiate, in the non-native condition, between clusters that are phonotactically legal for them (i.e., they occur in both languages) and those clusters unfamiliar to them (legal in Georgian only).

In the first instance, we confine the analyses to the clusters that exist in both languages. Note that the reference distributions of the audio models relative to which the Mahalanobis distance values were calculated contain all audio model data (as given in **Figure 3**); this way MDR values are comparable across the different analysis sections of this paper.

On the left of the boxplot in **Figure 5** we see the MDR values for the German speakers separately for the native and non-native conditions. There is a clear difference between the two conditions with the native condition being closer to the German audio model (negative

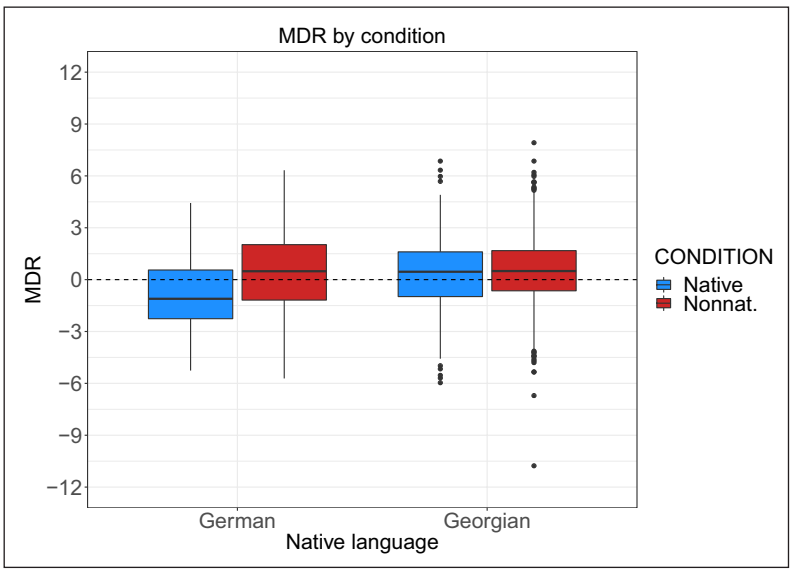

Figure 5: MDR values by speaker group and condition; only clusters existing in both languages are included. A negative MDR means that a given token is closer to the German audio model than to the Georgian audio model.

values) and the non-native condition being closer to the Georgian audio model (positive values). For the Georgian speakers, on the right-hand side of the graph, both medians are positive, indicating that the productions are overall closer to the Georgian audio model. By visual inspection there is no difference between conditions for the Georgian speakers, and both conditions are closer to the zero line compared to the German speakers (the zero line indicates equidistance to the two audio model distributions).

For statistical analysis, we considered a model with a 3-way interaction Native Language × Condition × Cluster; this model did not converge. We therefore designed a model that takes account of the two-way interactions and thus corrects for the effects of Cluster in the data but we do not evaluate cluster effects at this point of the analysis, where we are interested in overall adaptation effects between conditions. We consider cluster-specific effects in a separate analysis below. The statistical model had the dependent variable MDR with the fixed factors Condition (native, non-native), Native Language (German, Georgian), and Cluster (7 levels). Repetition was included as a fixed factor. The interactions included were Native Language by Condition, Cluster by Condition, and Cluster by Native Language. Random factor was random by-Condition intercept and slope for Participant. All effects (except for Repetition) and interactions were significant: Condition: $F(24.9) = 8.1$, $p = .009$; Native Language: $F(18.6) = 5.4$, $p = .03$; Cluster $(1177.68) = 21.6$, $p < .001$; Native Language × Condition: $F(18.6) = 7.5$, $p = .013$; Condition × Cluster: $F(1180.8) = 2.9$, $p = .007$; Native Language × Cluster: $F(1178.1) = 4.3$, $p < .001$. The important aspects of these results at this point in the analysis are that the two speaker groups differ significantly from each other overall, as well as the significant Condition × Native Language interaction. This significant interaction arises because the German speakers show a larger differentiation between the conditions compared to the Georgians.

In the preceding analyses, we asked how much a given token approximated the overall *distributions* of the audio models which the participants heard. The results revealed a significant difference between the two speaker groups. It is notable that the two speaker groups also differ in their relative distance to their native audio models. Therefore it is important to also evaluate the relative change between conditions for the two speaker groups independently of the audio model distributions.

We thus paired the data from the native and non-native conditions produced by each speaker in order to assess the degree of change each speaker exhibits between conditions without explicit reference to the audio model distributions. We correlated separately for the two speaker groups the productions from the native and the non-native conditions. Samples from the two conditions were paired as follows: For each speaker, we computed a by-cluster average for their native productions. Each token for a given speaker and given cluster produced in the non-native condition was then paired with that speaker's native average for that cluster. For instance, German speaker DE1's native /bl/ productions were averaged and each of DE1's non-native /bl/ productions was paired with their native /bl/ average. We then correlated these matched native and non-native production values across speakers within a given language group. If speakers did not change their production pattern for a given cluster between conditions, there should be a very strong correlation between their native and non-native productions. To the degree that they changed their production pattern between conditions, a relatively weaker correlation should hold.

Figure 6 gives the scatterplot for the normalized plateau lag measure. Both correlations were significant which is not surprising given we are correlating speakers' own productions across conditions. More interesting is here the marked difference in the strength of the correlation for the two speaker groups. For the German speakers, the correlation was weak with $r = .21$ ($p < .001$), reflecting what we have already observed in terms of the MDR ratios: German speakers change their production patterns between the native and non-native conditions. In contrast, for the Georgians the correlation was comparatively strong with $r = .65$ ($p < .001$) which again confirms that the Georgian speakers do not seem to adapt their consonant coarticulatory timing patterns between the native and non-native conditions.

We now investigate whether speakers react to finer grained differences between clusters within the audio model productions and show cluster-specific effects in their non-native productions. Observe, for instance, how in **Figure 3** the distance between /kl/ between the German and Georgian audio model productions is much larger than between the two /ʃp/ productions because within Georgian, /kl/ has a much larger onset lag value

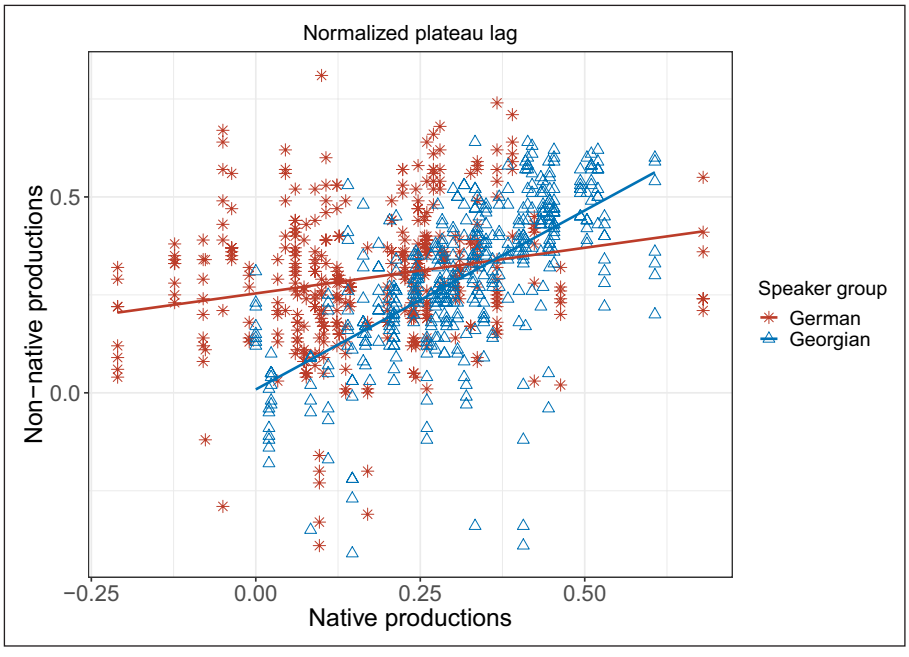

Figure 6: Scatterplot for normalized plateau lag pairing on a by-cluster basis each speaker's native and non-native productions, separately for the two speaker groups.

compared to /ʃp/. If German/Georgian speakers imitate these cluster specific differences, we should see a greater change in MDR value for, e.g., /kl/ than /ʃp/. We tested this for the German and Georgian data separately. The statistical model had the dependent variable MDR and the independent variables Condition (native, nonnative), Cluster (7 levels), and Repetition and a by-participant random intercept. Models with a random slope did not converge. If speakers' adaptation to the non-native audio model varies as a function of cluster, we expect a significant interaction. For the German speakers, Condition was significant ($F(549.04) = 15.9$; $p < .001$), as was Cluster ($F(549.48) = 6.4$, $p < .001$) and indeed the interaction ($F(548.65) = 4$, $p = .0006$). For the Georgian speakers, Condition was not significant ($F(642.46) = 2.6$, $p = .11$), but Cluster ($F(642.75) = 21$, $p < .001$) and the Interaction were ($F(642.08) = 3.5$, $p = .002$).

The significance of this interaction alone tells us that speakers' pronunciation changed to different degrees between conditions as a function of cluster, but it does not tell us about the direction of change. **Figure 7a** and **7b** illustrate for each cluster the average change per speaker between the native and the non-native condition. The arrows always point in the direction of change from native to non-native condition. The audio model productions for the clusters are also indicated in each panel. We can observe that the German speakers quite consistently changed their productions towards the Georgian audio model. The Georgian speakers, for whom there was overall no effect of condition, do show changes between conditions on a by-cluster level, hence the significant interaction. However, in contrast to the German speakers, the Georgian speakers' production behavior was not necessarily influenced by the German audio model. The arrows point in all directions, and there was generally less change for the Georgian speakers.

We next proceed to look at the role of the phonotactic status of the cluster, i.e., whether the cluster is legal in both languages or in Georgian only. Here we have to treat the German and Georgian data separately and begin with the Georgians. Recall that the German cluster inventory is a subset of the Georgian cluster inventory and hence all clusters are phonotactically legal for the Georgian participants. Yet since the Georgians all had some proficiency level in L2 German, and given that native patterns are known to drift towards an ambient L2 language (Chang, 2013; Sancier & Fowler, 1997), it does seem relevant to look for this speaker group at the native condition and differentiate clusters that exist in Georgian only versus those that exist in both languages. **Figure 8a** gives the MDR Distance values for all data recorded from the Georgian speakers for the native condition by cluster status (both languages, Georgian only).

Statistically, we compared the MDR values of the two groups in **Figure 8a** with fixed factors Cluster Status and Repetition and by-participant random intercept and slope for Cluster Status. Note that inclusion of Cluster as fixed factor was not possible here due to the collinearity between Cluster Status and Cluster. The effect of Cluster Status was not significant ($F(10.11) = 1.2$ and $p = .3$). We conclude that there is no evidence for a difference between the two cluster groups in the Georgian speakers' native productions in MDR values.

We now turn to the German participants, the main group of interest of this analysis. **Figure 8b** gives the MDR values as a function of Condition and Status. The boxplot suggests a three-way distinction for the German speakers. Since we have already seen above that the MDR values for the German speakers for the clusters that exist in both languages differ significantly from each other (significant interaction in **Figure 5**), we ran here a model comparing the two cluster groups in the non-native condition. The model had the dependent variable MDR and the fixed factors Cluster Status and Repetition, and a random intercept and slope for Cluster Status by Participant. Cluster Status was significant ($F(8.21) = 60.8$; $p < .001$).

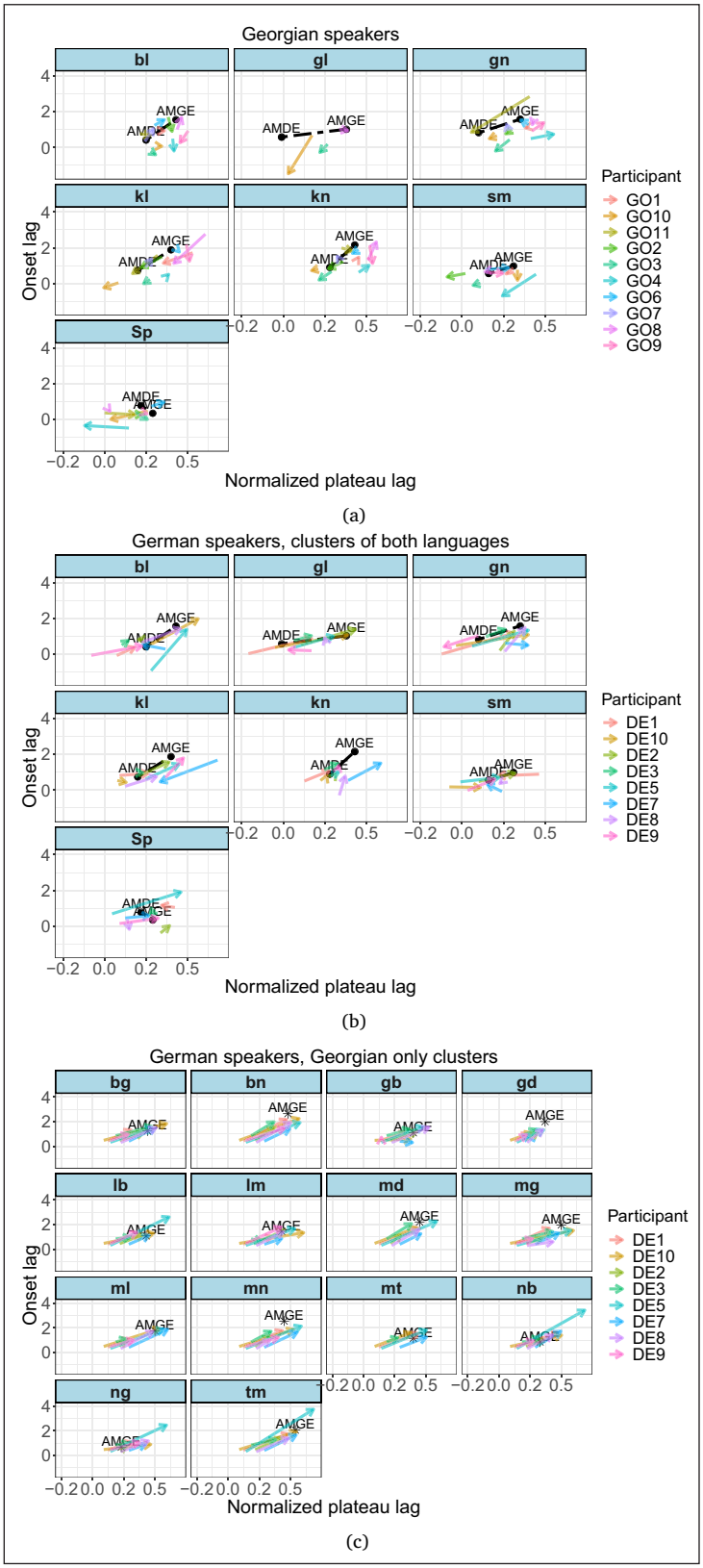

Figure 7: Average change for each speaker and cluster between the native and non-native conditions. In Figures **(a)** and **(b)**, the black dashed line indicates the distance between the two audio model productions. AMDE and AMGE stand for the German and Georgian audio model, respectively. <Sp> stands for /ʃp/. Figure **(c)** shows German speakers' productions of the clusters existing in Georgian only. In this case the native/starting point of the arrow is the mean of each speaker's native condition productions across clusters. The Georgian audio model productions are indicated by an asterisk and the letters AMGE.

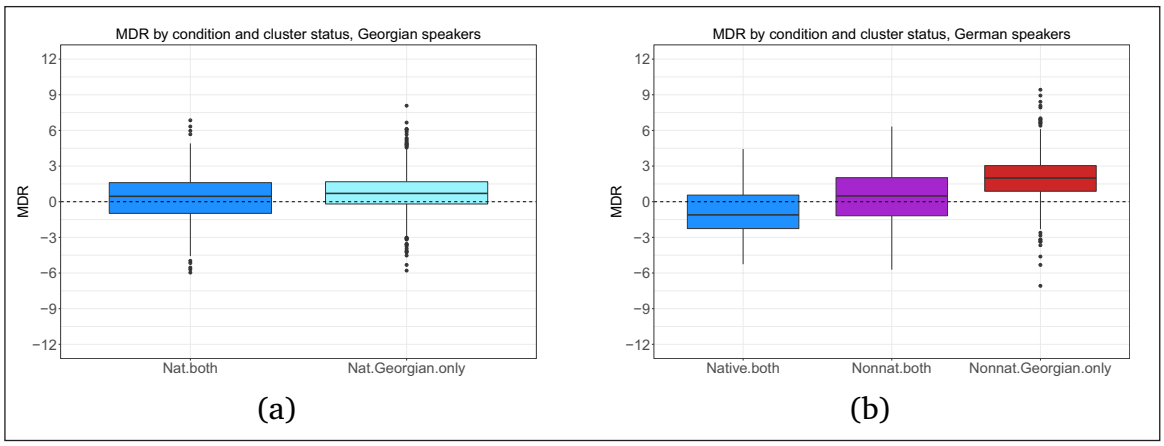

Figure 8: (a) MDR values of the Georgian speakers for the native condition by cluster status. **(b)** MDR values of the German speakers by condition and cluster status.

The three-way distinction we find for the German speakers is also visible in **Figure 7c**, which gives German speakers' productions of the clusters existing in Georgian only. Here, the native/starting point of the arrow is the mean of each speaker's native condition productions across clusters. We see overall longer arrows in **Figure 7c** compared to **Figure 7b**; the latter displays the clusters existing in both languages. This clearly illustrates the much more pronounced adaptation effect for clusters for which no native pattern acts as a constraining force on imitation.

Summarizing, we observed based on our landmark measures that the German speakers were close to their own audio model in the native condition and imitated in the nonnative condition the timing pattern of the Georgian audio model. While we saw imitative flexibility for all clusters for the German speakers, the degree of adaptation varied significantly as a function of the phonotactic status of the clusters in German. For clusters legal in German, speakers adopted an intermediate pattern between the Georgian and German reference pronunciations. For the clusters illegal in German, speakers showed the greatest distance to the German audio model. Importantly, for both conditions, degree of imitation can be predicted by audio model distance. Georgians, on the other hand, did not contrast between the native and nonnative conditions; the cluster-specific changes between conditions are not systematically in the direction of the German audio model. The relative greater distance the Georgians have to their native audio model may point to the Georgians imitating neither their native nor the German audio model, a point we shall return to in the Discussion.

It is possible that the reduction of coarticulatory timing patterns to a two-dimensional space fails to capture other aspects of the timing patterns which the Georgian participants may have zoomed into. While our lag measures are a convenient analysis tool to differentiate between the languages, speakers may well perceptually parse or control in their productions other parameters. To avoid incorrectly concluding that Georgian participants did not adapt their timing patterns at all, we now turn to a machine learning analysis that does not depend on the choice of specific landmark measures. This also allows us to investigate to which extent the languages and conditions are separable based on coarticulatory timing information at a global level.

3.2.3. Separating timing patterns by machine learning

Importantly, and problematically for our goal, hypotheses about timing patterns can be formulated in a high-dimensional space, with speakers potentially controlling the temporal relation between other landmarks than as described above. There is a myriad of logically

possible ways to quantify coarticulatory timing and there are indeed different proposals on the underlying control structures that give rise to any observed articulatory differences (see e.g., Tilsen, 2016). Hence it may be premature to conclude based on our analyses so far that Georgian participants do not differentiate between the native and nonnative conditions. We therefore use Support Vector Machines (SVMs), a machine learning tool (Bennet & Campbell, 2000), as an unbiased global approach to separating articulatory patterns between speaker groups and conditions. SVMs cannot always provide insight into the pattern they find, but they do allow us to find any pattern that separates data with minimal a priori assumptions about such patterns.

SVMs are a form of supervised learning and consist of finding an optimal separation between two categories. The separation is defined as an (n-1)-dimensional hyperplane fitted computationally through two sets of n-dimensional data points representing each category. If there are hyperplanes separating the two categories, the hyperplane furthest from the data points is used (i.e., in the middle of the two point clouds formed by the categories, not adjacent to one of those clouds). If the two categories overlap, the separation is still the one that separates most points of both categories. Fitted SVMs do not offer meaningful insight into the way the categories are separated, only into the practical separability. The main reason for this is the high dimensionality of the plane, which makes it impossible for the human eye/mind to make sense of the dividing hyperplane. Nevertheless, as our input only contains temporal information, any SVM that separates the two sets is assumed to separate the data based on timing patterns specifically.

Only clusters existing in both languages are included, as otherwise the SVM might separate based on timing patterns relative to the clusters instead of the native language of the speaker. For each consonant, the time points of six kinematic landmarks are used (see Section 2.5.2 in Methods). These time points are those of movement onset, peak velocity of constriction formation, target achievement, maximal constriction, release, and peak velocity of the release, for both the first and the second consonant. All time points were taken as relative to the velocity peak of the first consonant, which effectively removes that landmark as a potential contribution to a separation of the data as it is always zero. Hence, the data has eleven dimensions and a label for each native production.

Each dimension is normalized to the scale of 0–1, as otherwise larger numbers (i.e., the final time points of the second consonant) would have a disproportionate influence on the location of the separating hyperplane. Note that this normalization, performed for each dimension separately, might seem to obscure temporal order (e.g., a plateau onset of one data point might get a higher number than the corresponding plateau offset), but the relative order is preserved (e.g., data points with relatively long distances between onset and offset can still be distinguished from data points with shorter distances).[6]

[6] Rescaling each variable separately does technically not respect temporal order. Yet any information on temporal order is preserved under rescaling per variable and can serve to separate the data as follows: Taking as an example a two-dimensional plane with an X-range of {–100,100} and an Y-range of {0,3}, any line through the plane represents a SVM, a one-dimensional separation of the two-dimensional space into two categories. If we take the diagonal line through (–100,0) and (100,3), the point (0,0) is below this line. If we now rescale, by dividing X by 200 and adding 0.5 and by dividing Y by 3, the diagonal line now goes through (0,0) and (1,1) but the point (0,0) shifts to (–0.5,0) and is still below the line. This logic holds for any line and any point. However, given two separable point clouds, the line that best separates them (i.e., the one that is furthest from any of the points) is not the same after rescaling each variable separately, which is the rationale for rescaling. Returning to our actual data, in which dimensions are time points that have an order, it is possible to separate e.g., items for which C1 offset is after C2 onset from those for which this is not the case by drawing a diagonal line in the C1 offset-C2 onset plane through the origin. Any point below this line would have a C1 offset higher than its C2 onset (and thus have overlap), while any point above this line would have the opposite. Rescaling might lead to the number representing C1 offset to become lower than the number representing C2 onset even if was not before, but the data points of the other category would shift even more and the diagonal line can be rescaled with it. Crucially, this

To train the SVMs, we use only the productions of the native conditions, i.e., Georgian native speakers imitating the Georgian audio model and German native speakers imitating the German audio model. The productions of the audio model itself did not form part of the data set and one token had to be excluded as one time point had not been identified (see Methods). This native production subset used for training SVMs contains 302 data points, of which 130 are from German and 172 are from Georgian native speakers. The rest of the data set, never used for training, consists of the non-native productions, of which 355 are German native speaker imitations of Georgian and 383 are Georgian native speaker imitations of German.

First, parameters were explored with the use of the e1071 package for R (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2018), using the radial kernel. The best performance ($\gamma = .1$, cost $= 1$) had an error rate of .21, calculated with ten-fold cross-validation as provided by the tune.svm function of the e1071 package. Performance is clearly much higher than chance ($p < .001$). Given the satisfactory performance, we now explore in detail how SVMs perform on the native data set, as well as on the non-native productions, using these optimal parameters.

In order to increase the generalizability of our results, we trained 10,000 separate SVMs on subsets of the native productions as follows: For each iteration, the native productions were randomly divided into a training and test set, containing 90% and 10% of the data, respectively. With each SVM trained on a training set of native productions, both the test set of native productions and all non-native productions were classified as either Georgian or German native.

The model performance on the native production test set was as described below, with confidence intervals calculated using t-tests. Note that t-tests might not be completely accurate as observations might be correlated if models are very similar, but the narrow band of the intervals makes possible inaccuracies inconsequential.

The hit rate (defined as German productions correctly classified as such) was .748 (95% confidence interval [.746 – .751]). The false alarm rate (Georgian productions incorrectly classified as German) was .160 (95% confidence interval [.158 – .1620]). In terms of *d′*, which aggregates hit rate and false alarm rate, the model performance was consistently around 1.989 (95% confidence interval [1.968 – 2.010]). The behaviour of the models on the native data over iterations is shown in **Figure 9**.

It is possible that there are data points in the training set that have extreme values on one or more of the dimensions and if so, these could in a few iterations all end up and conspire to have a disproportionate effect on the location of the hyperplane. An SVM thus affected is expected to be overtrained and perform poorly on the test set. Therefore, we removed the SVMs that perform below the 95% confidence interval for *d′* (i.e, the poorest SVMs), which amounted to only 5 SVMs (so 9995 SVMs remained). With these SVMs, we inspect how non-native productions were classified.

Most productions were classified correctly, i.e., Germans imitating Georgian were mostly still classified as belonging to the German speaker group. However, the German participants did manage to fool the SVMs, i.e., have their productions classified as native Georgian productions, in 46.497% of the cases (95% confidence interval [46.438 – 46.556]%), while the Georgians did so less often: Their non-native productions were classified as native German in only 18.707% of the cases (95% confidence interval [18.677 – 18.738]%),

line would still exist. Hence, the temporal order of our tokens is no longer directly visible from the rescaled datapoints, but any aspect of a token related to temporal order is still accessible for separation of the data points.

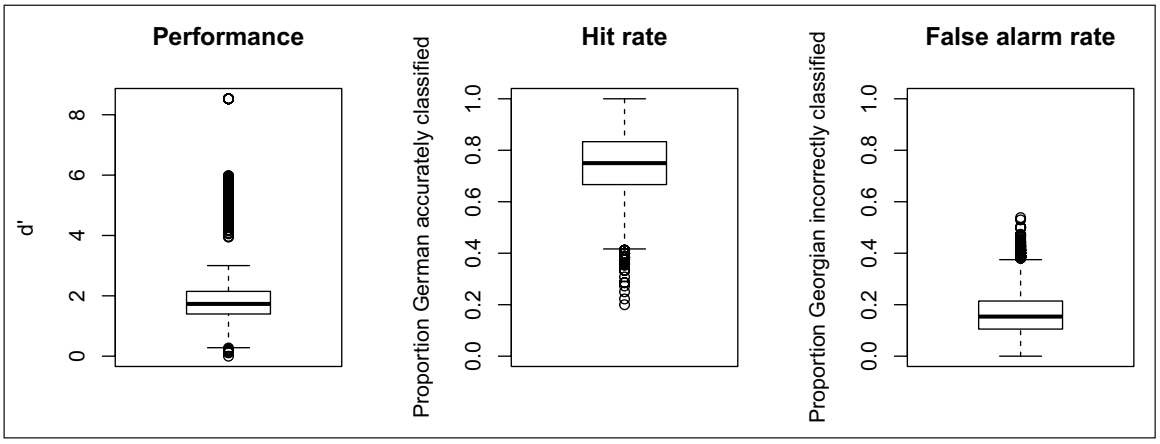

Figure 9: Performance (left), hit rate (middle), and false alarm rate (right) of SVMs on native data. Note that all data was classified as either German or Georgian. Each data point represents one iteration (one SVM) trained on 90% of the data and tested on the remaining 10%.

which is comparable to their native productions. The difference was significant (Mann-Whitney test, $W = 9.99 \times 10^7, p < .001$).

The separation is relatively stable with regard to the performance on the native and especially the non-native data. This stability indicates that the limited amount of the data did not impede successful generalization, which in turn suggests timing patterns were quite stable across the clusters and speakers used in this analysis.

In summary, it was possible to correctly identify the native language in the native condition in most cases with minimal a priori assumptions about the nature of the timing differences. In other words, the Georgian and German native productions were separable based on timing information, confirming the results obtained on the basis of the lag measures. The fact that the SVMs had a harder time classifying non-native productions for the German speakers once more confirms that these speakers adapted the timing of their productions to the audio stimuli. The fact that Georgian speakers' productions in the non-native condition were closest to their native productions in the most optimal separation of German and Georgian timing patterns strongly suggests that their level of adaptation was at best minimal.

3.2.4. Phonation profile of the transition

For our final analysis, we concentrate on the phonation profile of those clusters with what can be called an open transition period. The goal of the analyses presented in this section is to understand to which extent the phonation profile of the transition is part of the imitative process.[7] Given the size of our corpus and the variety of clusters included, and in order to keep the paper focused on the articulatory timing patterns, we confine ourselves to the clusters /bg, gb, gd, bl, gl/—the stop-stop clusters which exist in Georgian only and two clusters which exist in both languages with C1 being the same as one of the members in the stop-stop clusters. While a comprehensive acoustic analysis of the entire corpus is outside the scope of this paper, focusing on these two cluster groups will give an impression of the articulatory-acoustic relations and the degree to which the phonation profile of the transition is part of the adaptive process. It will be of particular interest to

[7] Another acoustic parameter in which we may expect to measure some degree of adaptation is the vowel. All stimuli contained the vowel /a/ which also differed in quality between the audio models. An F1/F2 formant analysis rendered no statistically significant adaptation effect in the vowel for either speaker group and is thus not reported in any further detail here.

which degree the German speakers will imitate the transition phonation profile of the Georgian audio model (which is markedly different from the German audio model; see e.g., **Figure 2**) for clusters legal in German. The stop-stop clusters analyzed here of course were part of the native condition only for the Georgians. An analysis of these clusters is included here mainly as a basis for comparison for the German speakers' non-native productions of these clusters.

In the first instance, we report transition duration for all clusters, as well as strength of voicing during C1 closure. In a second step we provide a phonation profile analysis of those clusters with a sufficiently long transition to allow for meaningful acoustic analysis (details below).

Recall that in our definition of transition, all tokens have a transition interval, which can be negative or positive (see Methods). A negative transition interval means that C2 achieved its target prior to the acoustic burst of C1. Such a case is illustrated in **Figure 10a**, which gives an example for a native German production with a negative transition of –30 ms; **Figure 10b** exemplifies a native Georgian production with a +64 ms transition. **Figure 11** gives the transition interval durations for all speaker groups and conditions as well as for the audio model speakers, separately for the two cluster groups.

In the German native condition for the stop-sonorant clusters, the median is fairly close to zero and we observe a pronounced difference between conditions. For the Georgians, we see that the median transition duration is slightly shorter in the non-native condition compared to the native condition. For the stop-stop clusters, the German median is very close to the Georgian audio models productions, consistent with the greater imitative flexibility our kinematic analyses revealed for the German speakers for clusters phonotactically illegal in German.

For the stop-sonorant group, we ran a mixed model on the durational values from the main experiment (i.e., excluding the audio models). We constructed a model in analogy to our MDR analysis above, with Duration as dependent variable and independent variables Native Language (German, Georgian), Condition (native, non-native), and Cluster and their two-way interactions. Repetition was also included as fixed factor. Speaker was random factor; a model with random slope

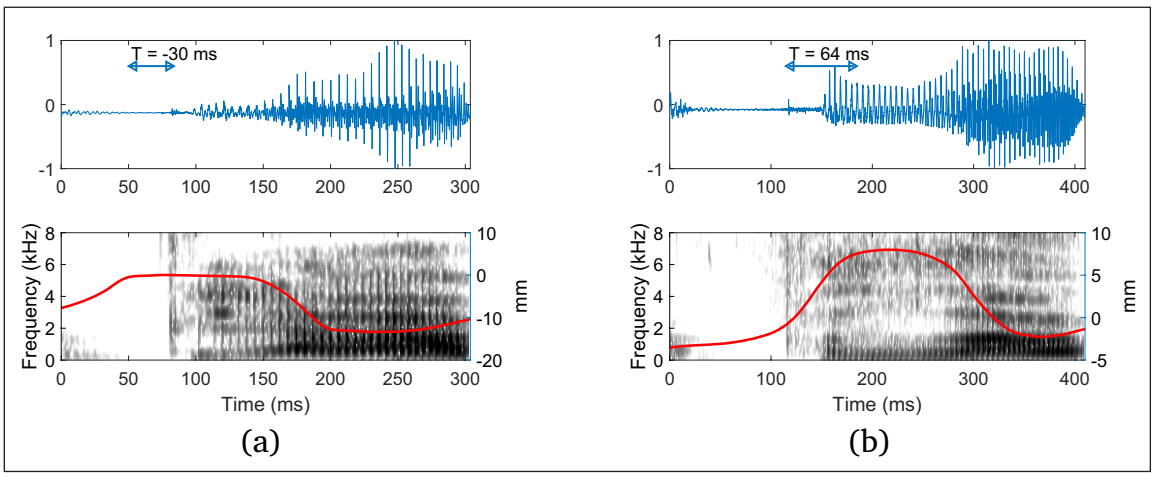

Figure 10: (a) German native production of /bl-/ with a transition interval of –30 ms. **(b)** Georgian native production of /bl-/ with a transition interval of +64 ms. In both figures, the vertical position time series of the tongue tip is superimposed on the sonagram for reference. The transition interval T is indicated by an arrow.

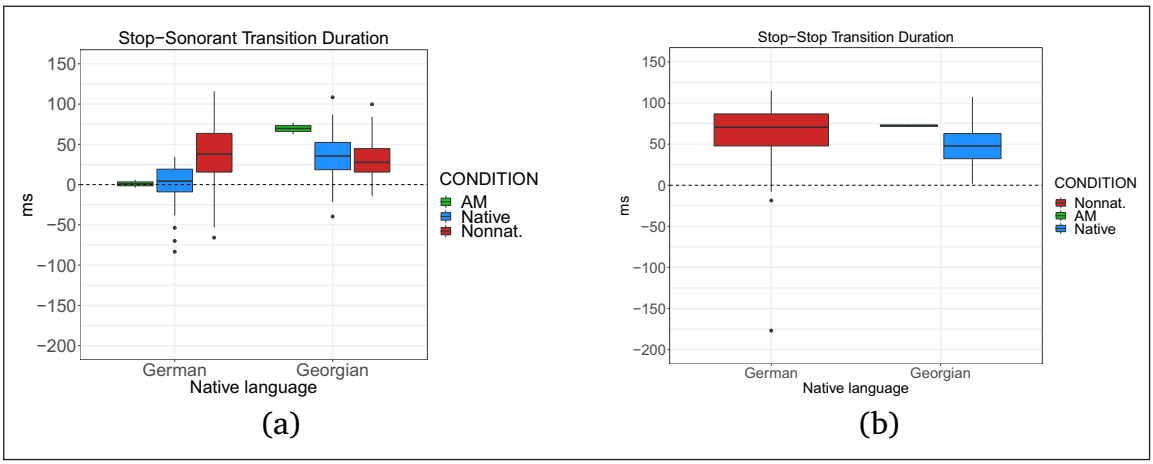

Figure 11: (a) Transition duration for the stop-sonorant group. **(b)** Transition duration for the stop-stop group. In both figures, the audio models are also shown. AM in the legend stands for audio model.

did not converge. Condition was significant ($F(215.8) = 15.8$, $p < .001$), as was the interaction between Native Language and Condition ($F(216.5) = 18.2$, $p < .001$). As in our previous analyses, the significant interaction underscores that German speakers change their productions to a larger degree between conditions compared to the Georgian participants.

Our next analysis concerns the amount of voicing during C1 closure and the phonation profile of the transition. Note that voicing during C1 closure cannot be interpreted within the context of adaptation, since all stop-initial audio model stimuli were cut at the burst. This analysis mainly serves as background information when it comes to comparing the phonation profile of the transition for the speaker groups, since differences in the phonation profile might be conditioned by stop voicing differences. Recall that we quantified the degree of voicing during the closure of C1 by taking the median of the normalized absolute amplitude of the signal during stop closure. Stronger voicing during closure is associated with higher amplitude values. For instance, for the token in **Figure 10a** this analysis rendered a value of .0014; the token in **Figure 10b** had a value of .0008. For reference, a strongly prevoiced token had a value of 0.026. The range of values across both cluster groups and all speakers was .0001 – .031.

Table 1 gives the median normalized signal amplitudes during C1 closure by condition, speaker- and cluster groups. The normalized amplitude values are very similar for the two speaker groups, consistent with what we had gleaned from the literature about the phonetic variability of phonologically voiced stops in both languages. A statistical model for the stop-sonorant group was run in analogy to the transition duration analysis. The main point of interest here is that there was no significant effect of Native Language ($F(18.4) < 1$, $p = .68$). This indicates that transition phonation profile differences between our speaker groups are unlikely to be primarily conditioned by language specific stop voicing effects.

We now move on to the phonation profile of the actual transition. For the following analyses, only tokens with transitions $> +10$ ms were taken into account (see Methods). **Table 2** gives the percent tokens for the two speaker groups as well as the audio models with transitions $> +10$ ms. Note that for the Germans the proportional amount of tokens produced with a $> +10$ ms transition almost doubled between the native and non-native

Table 1: Median with standard deviation in parentheses for the median normalized signal amplitude during C1 closure.

Consonant Group	Speaker Group	Native Condition	Non-native Condition
Stop-Son.	German	.0015 (.0036)	.0024 (.0057)
	Georgian	.0014 (.0087)	.0032 (.007)
Stop-Stop	German	n/a	.0014 (.0046)
	Georgian	.0011 (.0055)	n/a

Table 2: Percent tokens produced with a transition exceeding +10 ms.[8]

Consonant Group	Speaker Group	Native Condition	Non-native Condition	Audio Model
Stop-Son.	German	41%	78%	0%
	Georgian	81%	80%	100%
Stop-Stop	German	n/a	96%	n/a
	Georgian	97%	n/a	100%

condition for the stop-sonorant group, whereas for the Georgians there was no change between conditions.

For tokens with $> +10$ ms transitions, normalized zero crossing rate and amplitude envelope modulation ΔE were computed (see Methods). To illustrate the variety of phonation profiles that could be observed and exemplify the range of analysis values, we give examples for transitions from the stop-sonorant group in **Figure 12**, and for the stop-stop sequences in **Figure 13**. **Figure 14** gives the results for ΔE and the normalized zero crossing rate measures by speaker group.

We first turn to the stop-sonorant group. We observe a shallower amplitude modulation for the German native condition as indicated by the ΔE median being around zero. This contrasts with the pronounced amplitude modulation in the non-native condition. The Georgian speaker group is characterized by considerable amplitude modulation throughout. It is noteworthy that the variability pattern between the conditions reverses for the Germans and the Georgians. While the Georgians are noticeably variable in the native condition, the non-native condition has a markedly reduced variability in terms of the interquartile range. For the Germans, this effect is somewhat less, but in the opposite direction with the non-native condition having a larger interquartile range. We will return to this observation in the Discussion.

In terms of zero crossing rate, the Georgian compared to the German native productions have a higher degree of periodicity (lower zero crossing rate). Again we see the by now familiar pattern with the Germans changing their production between conditions to approximate the Georgian audio model values, in contrast to no change between conditions for the Georgian speakers.

[8] For the stop-sonorant consonant group, audio model percentages are out of a total of two tokens (one /gl/, one /bl/ production) per audio model. For the stop-stop consonant group, the Georgian audio model productions are out of a total of 3 tokens, one production each of /bg, gb, gd/. Audio model values are given for reference only.

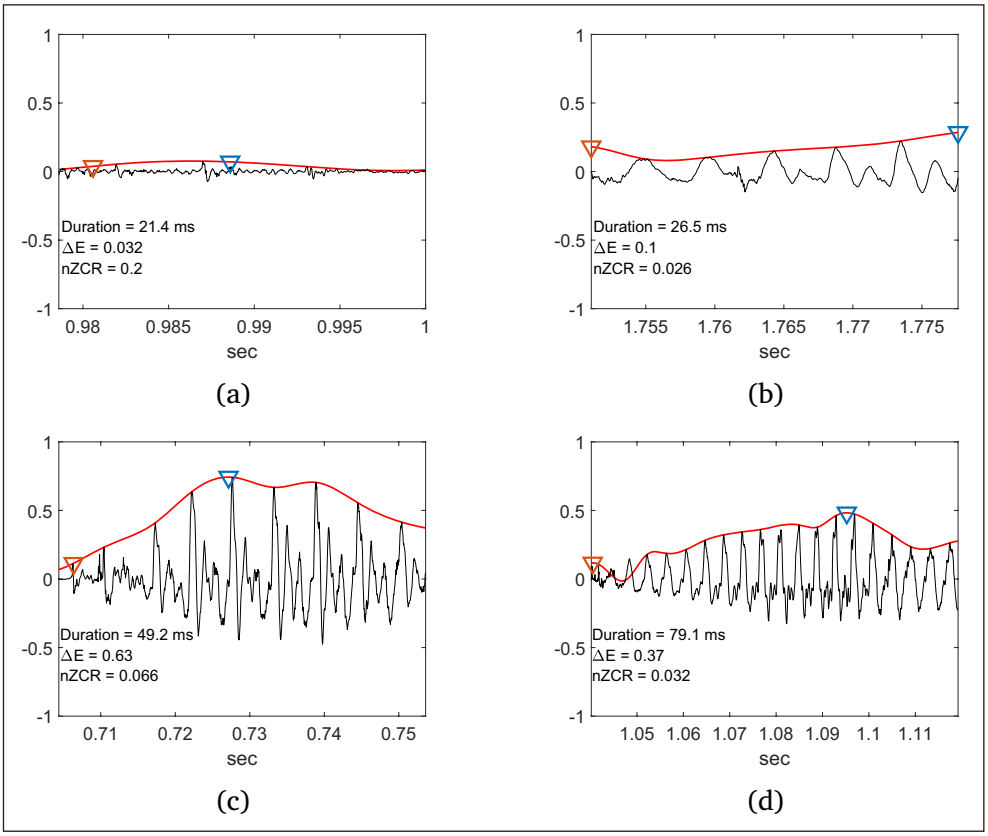

Figure 12: Oscillogram with amplitude envelope of transition intervals for stop-sonorant productions by German speakers in **(a, b)** and Georgian speakers **(c, d)**. On the left (a, c) are native productions, on the right non-native productions (b, d). (a) is from a /gl-/ production, all others are /bl-/. nZCR stands for normalized zero crossing rate. The triangles in each graph indicate E1, E2 which were used to calculate ΔE (E1 was the average of the envelope amplitude values of the initial 2 ms; see Methods).

Note that for reference, the audio models are part of **Figure 14**, but the audio model data were not included in the statistical analyses. To corroborate our visual observations statistically, we ran mixed models with the dependent variable being either ΔE or normalized zero crossing rate, in analogy to the stop voicing analysis above, i.e., independent factors were Condition, Native Language, Cluster, and all two-way interactions, as well as Repetition. There was a random intercept for Speaker. For ΔE, we again find a significant interaction between Native Language and Condition ($F(223.16) = 13.04$, $p = .0004$); the same holds for normalized zero crossing rate ($F(224.66) = 13.78$, $p = .0003$).

Turning to the stop-stop group (**Figure 14c, d**), we see that the Germans are very close to the Georgian audio model in their productions in terms of the transition envelope and normalized zero crossing rate.

Overall, the acoustic analyses underscore that the phonation profile of the transition is part of what the German speakers imitate. It is noticeable particularly for the stop-sonorant group—clusters which are part of the native German inventory—that the speakers produced phonation profiles very similar to the Georgian audio model, whereas for the MDR values, the productions were overall characterized by an intermediate pattern. This provides some evidence for the phonation profile of the transition being more than a by-product of a particular coarticulatory timing pattern in that it may be subject to aerodynamic control.

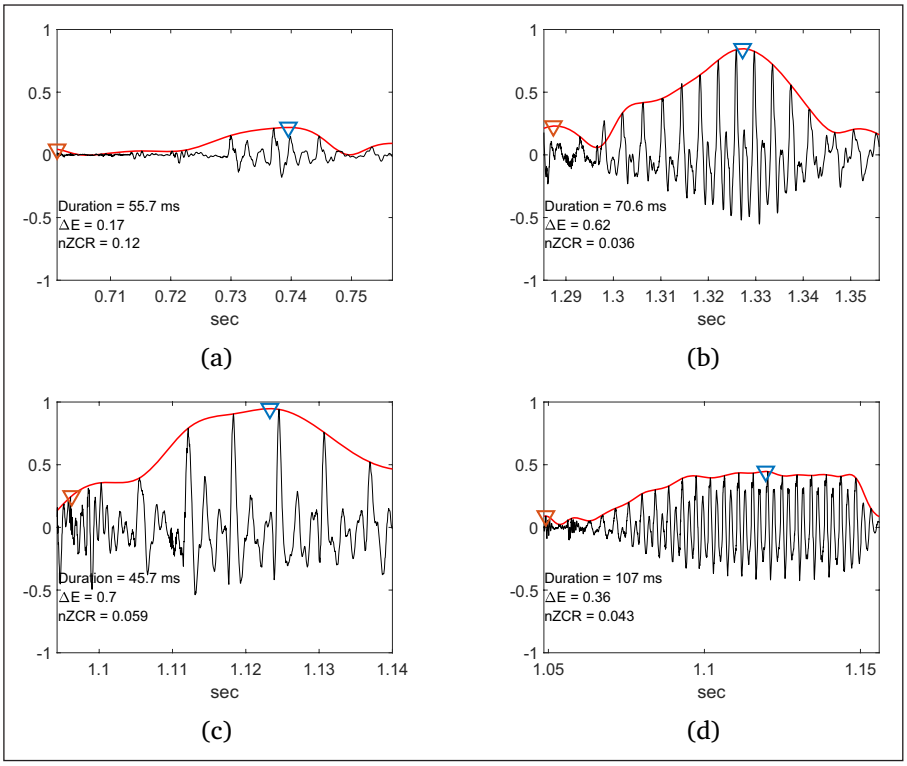

Figure 13: Oscillogram with amplitude envelope of transition intervals for stop-stop productions by German speakers **(a, b)** and Georgian speakers **(c, d)**. Stop-stop sequences are non-native for the Germans, native for the Georgians. nZCR stands for normalized zero crossing rate. The triangles in each graph indicate E1, E2 which were used to calculate ΔE (E1 was the average of envelope amplitude values of the initial 2 ms; see Methods). (a, c) are /gb/, (b) /gb/ and (d) /gd/.

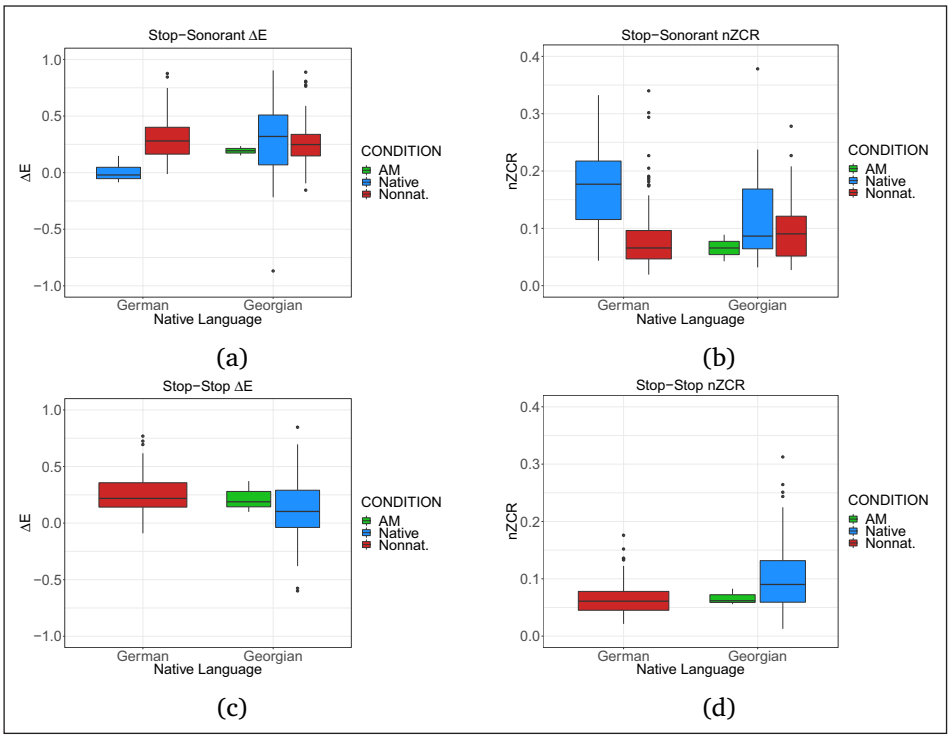

Figure 14: ΔE and normalized zero crossing rate (nZCR) for the stop-sonorant group in **(a, b)** and the stop-stop group in **(c, d)**. AM in the legend stands for audio model. The German audio model is not included because she did not produce any transitions > +10 ms.

4. Discussion

The goal of our study was to broaden our understanding of how coarticulatory timing patterns are adapted in an imitation task for both known and unfamiliar (phonotactically illegal) clusters. To that effect, we contrasted German and Georgian, two languages with different consonant-consonant timing patterns in CCV clusters. While Georgian participants were familiar with all clusters presented to them, for the German participants, we were able to probe the plasticity of coarticulatory timing patterns for both familiar and unfamiliar clusters.

Our results for the German speakers extend the existing literature on the imitation of fine phonetic detail by showing that speakers are able to adjust their native C-C coarticulatory timing pattern towards the pattern of another language even in a native language setting (German carrier phrase). Nevertheless, German participants imitated the Georgian C-C timing patterns for clusters familiar to them only to a certain degree. Hence, German participants showed signs of both stability and flexibility by producing clusters with an intermediate pattern between the two audio models. This was also underscored by the SVM classification which performed close to chance level for the German non-native condition (recall that this analysis included only clusters existing in both languages). For clusters existing in Georgian only, speakers moved to a coarticulatory timing pattern more similar to the Georgian audio model and further away from their own native timing pattern.

We could further show that German speakers imitated cluster specific coarticulatory timing patterns, since the MDR ratio differed significantly as a function of distance between the audio models (see **Figure 7**), crucially, for both cluster groups—those familiar to German speakers and those phonotactically illegal in German. Zsiga (2003), on the other hand, had found in the context of an L2 study that L1 English speakers of L2 Russian adopted a low overlap pattern across the board similar to Russian, but not quite like Russian because they failed to produce Russian-like fine-grained differences between clusters. She therefore argued that English speakers' productions were governed by the emergence of the unmarked rather than being an adoption of the Russian L1 pattern.

There are many differences between Zsiga's study and ours, but several of them seem particularly relevant here. Importantly, Zsiga's participants performed a reading, not an imitation task and thus did not react immediately to auditory input as our speakers did. Our study suggests that in principle there can be a differentiated by-cluster imitation and hence by implication, learning, of unfamiliar, cluster-specific consonant overlap patterns. Under which circumstances these by-cluster patterns may be stabilized in the context of L2 acquisition is not addressed by our present work. Secondly, Zsiga's English speakers may have partially adopted Russian cluster-specific timing differences but these may have been unrecoverable by acoustic analysis, especially since she focused on stop-stop sequences. Lastly, in our imitation study the unfamiliar coarticulatory timing pattern was produced as part of a native carrier phrase. Even clusters phonotactically legal and produced in a German carrier phrase showed a significant change in coarticulatory timing pattern in the imitation task, giving evidence for the flexibility of coarticulatory timing patterns in a native language setting. Whether phonotactically known sequences would be adapted to an unfamiliar coarticulatory timing pattern to a larger degree when produced in an L2 setting has to be addressed by a different experiment. Preliminary work by Kwon and Chitoran (2018) suggests such a scenario.

Our German participants' behavior contrasted starkly with the Georgian speakers who stayed with their native pattern throughout the experiment. This observation was underscored by several lines of analyses: The evaluation of lag measures relative

to the audio models, the degree of correlation between speaker's native and non-native productions, and the SVM results. The acoustic results further confirmed the asymmetry between the two speaker groups. Georgian speakers changed to a much lesser degree in the non-native condition compared to the Germans and not consistently in the direction of the German audio model, as particularly evident in **Figure 7**.

Asymmetric adaptation patterns have been reported in the literature before, notably Zsiga (2003). She attributed the behavior of the Russian speakers (who retained their native coarticulation pattern in L2 English) to a word integrity effect which surfaces in L2 at word boundaries, possibly part of a general functional effect in L2 production which by hypothesis maximizes cue recoverability. Since Russian already phonologically prioritizes a perceptually optimal production pattern (though see Pouplier et al., 2017), the surfacing of the unmarked cannot be differentiated from native language transfer. Our results are generally consonant with such a markedness hypothesis: Non-native patterns should be readily produced if unmarked or favored by phonetic biases. If Georgian, similar to Russian, has phonologized a pattern that prioritizes perceptual cue recoverability in consonant sequences over other aspects of speech production (Chitoran et al., 2002), 'unmarked non-native' productions could at least on the surface be arguably indistinguishable from native productions.

Yet there are several other ways to interpret our results. First of all, the fact that Germans adapted their native clusters in a native carrier sentence towards the Georgian audio model seems problematic for the view that non-native production necessarily prioritizes a certain type of cue recoverability, or that heightened functional pressure will coerce a change in articulatory routines. In our study, speakers were presented with unfamiliar coarticulatory timing patterns, but the imitation task was part of their native language performance, as they had to integrate the unfamiliar pattern into a native language carrier phrase. The clusters existing in both languages are arguably a priori optimally recoverable for German speakers within a German carrier phrase. Moreover there are no particular recoverability principles at stake for say, /bl/, which is probably perceptually one of the 'best' clusters (e.g., Henke, Kaisse, & Wright, 2012). Since the non-native coarticulatory timing pattern is produced as part of a German carrier phrase, it seems questionable to assume that the imitation we measure is driven by general principles of perceptual recovery.

In this context, it is important to also consider the results of our acoustic analysis for two of the stop-sonorant clusters. For the native clusters /bl, gl/, we saw that the German speakers converged towards the Georgian audio model in terms of normalized zero crossing rate and ΔE, and almost twice the number of tokens had a transition $> +10$ ms compared to the native condition. These non-native productions were in stark contrast to the native productions which were characterized by a high zero crossing rate and a flat amplitude envelope. This means that even though their production only changed to an intermediate degree in terms of the MDR value, the acoustics changed quite dramatically. Controlling the aerodynamics and hence phonation profile of the transition period may in fact be part and parcel of both native and non-native cluster production. Speculatively this may imply that the transition is acoustically not purely determined by context but may be subject to aerodynamic control. This clearly is a topic for future research. In the case of our experiment, it seems that the German speakers moved to the Georgian-like phonation profile pattern as part of faithful imitation, independently of functional pressures. Our analyses overall imply that open transitions can be observed not only in the context of gestural mistiming of phonotactically illegal clusters, but they may also emerge as part of a controlled imitation of an unfamiliar coarticulatory timing pattern. Speakers were able to flexibly extend the coarticulatory timing and phonation pattern part for the clusters/timing patterns part of their native grammar.

Overall, we would like to further suggest that we see quite clearly for the German speakers the constraining force of existing coordination patterns in the non-native condition, similarly to the way phonological contrast has been shown to constrain adaptation. Speakers adapt to some degree, but much less compared to the clusters phonotactically unfamiliar to them. This would also describe the results for the Georgian speakers: The German clusters are a subset of the Georgian cluster inventory, hence imitative flexibility is limited by the availability of a native pattern. Since all clusters produced in the experiment are part of the Georgian phonotactic inventory, we may see this constraining force for the Georgians speakers throughout the experiment.

This brings us to the question whether the imitation of the unfamiliar clusters by the German speakers is really a more flexible imitation of the Georgian audio model or whether the MDR values for those clusters arise from a categorically different production such as resyllabification or vowel insertion. In terms of the latter, it may be the case that some of the productions have an epenthetic vowel instead of an open transition, but we do not have a way of quantifying this in an objective fashion. The duration of the transitions cannot be used as an indicator here, since they may simply approximate the Georgian audio model. Neither can an emergence of formant structure be used in a straightforward fashion—an open transition may show formant structure when there is glottal vibration while conversely a very short or partially devoiced vowel can have very weak/no higher harmonic structure. Note that in none of the continuous measures we have applied to the data was there evidence for a categorically different behavior according to cluster status. Particularly the analysis of cluster-specific effects (**Figure 7**) underscored that the distance between audio model productions predicted the degree of adaptation for both native and non-native clusters.

As to the resyllabification question, word-final devoicing can provide some relevant cues: German has word-final devoicing which may be incomplete in terms of effects on preceding vowel duration, but not in terms of stop prevoicing (Nicenboim, Roettger, & Vasishth, 2018; Piroth & Janker, 2004); devoiced as well as fortis word-final stops are often aspirated (Kohler, 1995). At least for the stop-stop clusters, the high periodicity of the transition in **Figure 14** provides circumstantial evidence against the production being C^h#C with an aspirated word-final stop rather than a CCV production. In **Table 1**, there is no difference in C1 voicing in our measure between the Germans' non-native stop-stop productions and the native stop-sonorant productions. If C1 in the stop-stop productions were systematically resyllabified as a coda, we should see the effects of word-final devoicing in the normalized zero crossing rate. Lastly, for some of the productions we informally observed a glottal stop being inserted between carrier phrase and target word presumably as a word boundary marker which also speaks against resyllabification being the dominant factor conditioning the results.

A puzzling result of our study is the complete absence of imitation by the Georgian speaker group for any of our measures. We thereby suspect (although we have no way of testing this with the data we have) that the Georgian speakers do not imitate the Georgian audio model either, which may be why their productions are further from their own audio model compared to the Germans in **Figure 5**. We have to acknowledge here that the behavior of the Georgians may be related to them being part of an expat community and having some (limited to advanced) knowledge of German. It may be that their native productions were already characterized by an intermediate coarticulatory timing pattern with which they produced the stimuli in the native Georgian and also the nonnative German condition, which of course we would not be able to see in our experiment. (A native Georgian speaker informally listened to some of the recordings and did not find them to be unusual, but of course that does not mean that there were not subtle effects of native language attrition

(Sancier & Fowler, 1997), and this is a limitation of our current study. But native language attrition does not explain why the participants do not seem to imitate the German audio model, nor would we expect the Georgian audio model to be exempt from native language attrition. Note that her productions are quite distinct from the German productions and cannot meaningfully be analyzed in terms of a drift towards German.

A to us more promising aspect to consider in this context is the role of attention in two respects. For one, there is a much larger spread of overlap values in the Georgian audio model data than there is for the German audio model data. The German participants were exposed in the non-native condition to a much wider range of timing patterns, possibly directing their attention more towards this variation in the non-native stimuli. The presence of phonotactically illegal clusters in particular may have focused their attention much more on the phonetic properties of the stimuli and their own pronunciation of what they heard. Both speaker groups were alerted to the fact that they would hear stimuli from a different language in the non-native condition which may sound unfamiliar. This was truly the case for the German participants, but the Georgian speakers reported generally recognizing the non-native stimuli as German when informally asked after the recording session. Georgian speakers may have paid less attention to the phonetic detail of the stimuli once they realized that they were produced by a German speaker. Such a scenario would be in line with the observation by Nye and Fowler (2003) who found that the amount of language knowledge available mediates imitation. Nye and Fowler (2003) systematically varied the order of approximation of their stimuli to English phonetic sequences, and hence by implication the amount of language knowledge imitators could draw upon. The more the stimuli conformed to English the less closely the stimuli were imitated. This does not, however, explain the difference in approximation of our two speaker groups to their audio models.

But there may also be another, deeper role of attention in our results. Strange (2011) highlighted the role of language-specific cue weighting routines in second language learning. Selective attention to language-specific cues varies, according to Strange, with experience and form part of native speaker knowledge. This means that selective attention routines are language specific, highly overlearned, and automatic (they may also differ between individuals within a speech community, see recently Beddor, 2009; Coetzee, Beddor, Shedden, Styler, & Wissing, 2018). For L2 perception, these automatic attention routines have to be recalibrated (see also Chang, 2018; Dmitrieva, 2019). In the context of our present experiment, we would like to speculate that the differences in vowel reduction between German and Georgian may implicate different cue weighting patterns for coarticulatory timing in C-C onset clusters. Recall that Georgian does not have vowel reduction and the Georgian vowel inventory does not have a central mid vowel (Shosted & Chikovani, 2006). Germans, in contrast have ample experience with many different degrees of vowel reduction (Helgason & Kohler, 1996; Jannedy, 1994). This may mean that for clusters Germans may have quite acute attentional tuning to the kind of coarticulatory timing that we investigated. For Georgians, the temporal cues that our experiment quite narrowly zoomed in on are possibly less important than signal modulation of the transition. Circumstantial evidence for the importance of signal modulation of the transition may be provided by the variability of the amplitude envelope modulation we saw in **Figure 14** for the Georgian native condition, which may be indicative of systematic cue trading relations. This will be an interesting avenue for future research.

Finally we would like to point out that we observed some cases of surprisingly high overlap for the Georgian audio model productions. While Georgian productions can, to the extent that this is actually known, be characterized by a low overlap pattern and long positive transition intervals, it seems that relatively higher overlap productions are, at

least in certain circumstances, a viable production pattern in Georgian, too. Whether we have more variability in the audio model data for the Georgian audio model because we have more data from her (21 clusters as opposed to 7 from the German audio model) or whether Georgian is more variable in overlap patterns cannot really be assessed, since the confound that the German cluster inventory is a subset of the Georgian inventory is absolute. Beyond the study by Chitoran et al. (2002), which reports articulatory data from two speakers, very little is actually known about how Georgian clusters are articulated. Hermes et al. (2017) have raised the possibility that languages differ in their variability in cluster production. They argue that consonant cluster production in Polish is more variable than Tashlhiyt Berber: In Polish, but not Tashlhiyt Berber, the segmental composition of a consonant sequence conditions variability in coarticulatory timing. (Note though that in their study the recorded material between the two languages was not the same. Thus language, cluster composition, and syllable affiliation differences were confounded and it is hard to pinpoint the source of the observed effects.) A similar observation was made by Zsiga (2000) for English and Russian. Given further the literature on differences in the extent of vowel-to-vowel coarticulation cited in the Introduction, it seems plausible that languages may not only differ in their consonant timing as such, but also in the range of coarticulatory timing patterns they exhibit. Such differences are likely to impact speakers' flexibility in an imitation task, either in terms of their perceptual acuity for timing differences and/or their production flexibility. To our knowledge, next to no data exist on this issue and we see this as an important avenue for research, as increasing our understanding of how and in which direction coarticulation patterns may change may ultimately help render insights into how phonotactic patterns may be acquired, and how through the dynamics of language use, they may emerge and dissolve (Chitoran & Iskarous, 2008; Easterday, 2019).

Acknowledgements

Work supported by ANR-DFG grant PATHS, PO1269-31 to Ioana Chitoran and Marianne Pouplier and by the ERC under the EU's 7th Framework Programme (FP/2007–2013)/ Grant Agreement n. 283349-SCSPL. Thank you to Harim Kwon, Manfred Pastätter, and all members of the PATHS project for insightful discussions. Chris Carignan and Raphael Winkelmann provided very helpful advice for the phonation profile analysis. Our student RAs Susanne Waltl, Anna Ratzinger, and Dustin Young helped with data recording and segmentation.

Competing Interests

The authors have no competing interests to declare.

References

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. DOI: https://doi.org/10.1515/lp-2014-0006

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, *5*, 123–150. DOI: https://doi.org/10.1515/lp-2014-0006

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821. DOI: https://doi.org/10.1353/lan.0.0165

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. G., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*(4), 931–968. DOI: https://doi.org/10.1353/lan.2018.0051

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*, 591–627. DOI: https://doi.org/10.1006/jpho.2002.0177

Bennet, K. T., & Campbell, C. (2000). Support Vector Machines: Hype or hallelujah? *SIGKDD Explorations*, *2*(2), 1–13. DOI: https://doi.org/10.1145/380995.380999

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*(14), 5321–5325. DOI: https://doi.org/10.1073/pnas.0801469105

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*(3), 591–630. DOI: https://doi.org/10.1016/j.cognition.2006.05.015

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* [Computer Program]. Retrieved from https://www.praat.org

Bombien, L., & Hoole, P. (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. DOI: https://doi.org/10.1121/1.4807510

Boyce, S. E. (1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America*, *88*(6), 2584–2595. DOI: https://doi.org/10.1121/1.400349

Broselow, E., Chen, S.-I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*, 261–280. DOI: https://doi.org/10.1017/S0272263198002071

Butskhrikidze, M. (2002). *The Consonant Phonotactics of Georgian*. Utrecht: LOT Dissertation Series.

Byrd, D., & Tan, C. (1996). Saying consonant clusters quickly. *Journal of Phonetics*, *24*(2), 263–282. DOI: https://doi.org/10.1006/jpho.1996.0014

Carignan, C. (2018). Using naïve listener imitations of native speaker productions to investigate mechanisms of listener-based sound change. *Journal of Laboratory Phonology*, *9*(1), 1–31. DOI: https://doi.org/10.5334/labphon.136

Catford, J. (1985). 'Rest' and 'open transition' in a systemic phonology of English. In J. D. Benson & W. S. Greaves (Eds.), *Systemic Perspectives on Discourse. Vol 1: Selected Theoretical Papers from the Ninth International Systemic Workshop* (Vol. 1, pp. 333–349). Norwood, NJ: Ablex.

Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition*, *22*(1), 1–26. DOI: https://doi.org/10.1017/S0272263100001017

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 520–533. DOI: https://doi.org/10.1016/j.wocn.2013.09.006

Chang, C. B. (2018). Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics*, *68*, 85–102. DOI: https://doi.org/10.1016/j.wocn.2018.03.003

Chitoran, I. (1998). Georgian harmonic clusters: Phonetic cues to phonological representation. *Phonology*, *15*(2), 121–141. DOI: https://doi.org/10.1017/S0952675798003558

Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven, T. Rietveld & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 419–448). Berlin: Mouton de Gruyter.

Chitoran, I., & Iskarous, K. (2008). Acoustic evidence for high vowel devoicing in Lezgi. *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg, 93–96.

Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics, 4*, 337–351. DOI: https://doi.org/10.1016/S0095-4470(19)31260-4

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–216. DOI: https://doi.org/10.1016/j.wocn.2017.09.009

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678. DOI: https://doi.org/10.1121/1.1810292

Daland, R., Oh, M., & Davidson, L. (2018). On the relation between speech perception and loanword adaptation. *Natural Language and Linguistic Theory*. DOI: https://doi.org/10.1007/s11049-018-9423-2

Davidson, L. (2005). Adressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics*, *19*(6/7), 619–633. DOI: https://doi.org/10.1080/02699200500114077

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*(1), 104–137. DOI: https://doi.org/10.1016/j.wocn.2005.03.004

Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, *53*(8), 1042–1058. DOI: https://doi.org/10.1016/j.specom.2011.05.010

Davidson, L., Martin, S., & Wilson, C. (2015). Stabilizing the production of nonnative consonant clusters with acoustic variability. *Journal of the Acoustical Society of America*, *137*(2), 856–872. DOI: https://doi.org/10.1121/1.4906264

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*(145–173). DOI: https://doi.org/10.1159/000107914

D'Imperio, M.-P., Cavone, M., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*. DOI: https://doi.org/10.3389/fpsyg.2014.01226

Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics*, *73*, 128–143. DOI: https://doi.org/10.1016/j.wocn.2018.12.008

Easterday, S. (2019). Highly Complex Syllable Structure: A Typological and Diachronic Study. *Language Science Press*. DOI: https://doi.org/10.5281/zenodo.3268721

Eckman, F. (2008). Typological markedness and second language phonology. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 95–115). Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/sibil.36.06eck

Endresen, R. (1991). *Fonetikk og Fonologi*. Oslo: Univesitetsforlaget.

Farnetani, E., & Recasens, D. (2010). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd. ed.) (pp. 316–352). Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444317251

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65. DOI: https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human Communication and Its Disorders: A review* (pp. 244–401). Norwood, NJ: Ablex.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76*(3), 708–721. DOI: https://doi.org/10.1121/1.391257

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Effects of age of second language learning on the production of English consonants. *Speech Communication, 16*, 1–26. DOI: https://doi.org/10.1016/0167-6393(94)00044-B

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, *20*, 269–337. DOI: https://doi.org/10.1023/A:1014942312445

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. DOI: https://doi.org/10.1037/0033-295X.105.2.251

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition, 107*(3), 1155–1164. DOI: https://doi.org/10.1016/j.cognition.2007.11.009

Gow, D. W. J. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 163–179. DOI: https://doi.org/10.1037/0096-1523.28.1.163

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, *23*(3), 387–429. DOI: https://doi.org/10.1017/S0952675706000996

Hallé, P., & Best, C. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop + /l/ clusters. *Journal of the Acoustical Society of America, 121*, 2899–2914. DOI: https://doi.org/10.1121/1.2534656

Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). The processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 592–608. DOI: https://doi.org/10.1037/0096-1523.24.2.592

Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science, 10*(4), 1–21. DOI: https://doi.org/10.1111/tops.12329

Helgason, P., & Kohler, K. J. (1996). Vowel deletion in the Kiel corpus of spontaneous speech. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, *30*, 115–158.

Henke, E., Kaisse, E. M., & Wright, R. (2012). Is the sonority sequencing principle an epiphenomenon? In S. Parker (Ed.), *The Sonority Controversy* (pp. 65–100). Berlin: Walter de Gruyter.

Hermes, A., Mücke, D., & Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics*, *64*, 127–144. DOI: https://doi.org/10.1016/j.wocn.2017.05.004

Hirose, H., & Gay, T. (1972). The activity of the intrinsic laryngeal muscles in voicing control. An electromyographic study. *Phonetica, 25*(3), 140–164. DOI: https://doi.org/10.1159/000259378

Hoole, P. (2006). *Experimental studies of laryngeal articulation* [Habilitationsschrift, Ludwig-Maximilians Universität München, Germany]. Retrieved from http://www.phonetik.uni-muenchen.de/~hoole/pdf/habilpgg_chap_all.pdf

Jannedy, S. (1994). Rate effects on German unstressed syllables. *Working Papers in Linguistics. Papers from the Linguistics Laboratory Ohio State University*, *44*, 105–124.

Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents in German*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/sfsl.44

Kochetov, A., Pouplier, M., & Son, M. (2007). Cross-language differences in overlap and assimilation patterns in Korean and Russian. In *Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken* (pp. 1361–1364).

Kochetov, A., & So, C. K. (2007). Place assimilation and phonetic grounding: A cross-linguistic perceptual study. *Phonology*, *24*(3), 397–432. DOI: https://doi.org/10.1017/S0952675707001273

Kohler, K. (1995). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Kwon, H., & Chitoran, I. (2018). The adaptation of native clusters with non-native phonetic patterns is task-dependent. *Poster presented at the Representing Phonotactics Workshop*, 23 June 2018. Retrieved from http://labphon16.labphon.org/se-03-program.html

Lentz, T. O., & Kager, R. W. J. (2015). Categorical phonotactic knowledge filters second language input, but probabilistic phonotactic knowledge can still be acquired. *Language and Speech*, *58*(3), 387–413. DOI: https://doi.org/10.1177/0023830914559572

Lubker, J., & Gay, T. (1982). Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America*, *71*(4), 437–448. DOI: https://doi.org/10.1121/1.387447

Ma, L., Perrier, P., & Dang, J. (2015). Strength of syllabic influences on articulation in Mandarin Chinese and French: Insights from a motor control approach. *Journal of Phonetics*, *53*, 101–124. DOI: https://doi.org/10.1016/j.wocn.2015.09.005

Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, *88*(3), 1286–1298. DOI: https://doi.org/10.1121/1.399705

Marin, S. (2013). Organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics*, *41*, 211–227. DOI: https://doi.org/10.1016/j.wocn.2013.02.001

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception and Psychophysics*, *34*(4), 338–348. DOI: https://doi.org/10.3758/BF03203046

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21–46. DOI: https://doi.org/10.1006/jmla.1998.2568

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2018). e1071: Misc functions of the Department of Statistics, Probability theory group (formerly: E1071), TU Wien [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=e1071 (R package version 1.7-0).

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*(2), 338–352. DOI: https://doi.org/10.1121/1.1907526

Mok, P. (2010). Language-specific realizations of syllable structure and vowel-to-vowel coarticulation. *Journal of the Acoustical Society of America*, *128*(3), 1346–1356. DOI: https://doi.org/10.1121/1.3466859

Mok, P. (2012). Does vowel inventory density affect vowel-to-vowel coarticulation? *Language and Speech*, *56*(2), 191–209. DOI: https://doi.org/10.1177/0023830912443948

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, *84*, 55–71. DOI: https://doi.org/10.1016/S0010-0277(02)00014-8

Nicenboim, B., Roettger, T. B., & Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in German. *Journal of Phonetics*, *70*, 39–55. DOI: https://doi.org/10.1016/j.wocn.2018.06.001

Nielsen, K. (2008). *Word-level and Feature-level Effects in Phonetic Imitation* (Unpublished doctoral dissertation). University of California at Los Angeles.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142. DOI: https://doi.org/10.1016/j.wocn.2010.12.007

Nye, P. W., & Fowler, C. A. (2003). Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, *31*(1), 63–79. DOI: https://doi.org/10.1016/S0095-4470(02)00072-4

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*(1), 151–168. DOI: https://doi.org/10.1121/1.1909864

Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America*, *114*(3), 1582–1599. DOI: https://doi.org/10.1121/1.1587737

Piroth, H. G., & Janker, P. M. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics*, *32*(1), 81–109. DOI: https://doi.org/10.1016/S0095-4470(03)00008-1

Polivanov, E. (1931). La perception des sons d'une langue étrangère [Perception of sounds of a foreign language]. In *Travaux du Cercle Linguistique de Prague: Vol. 4. Réunion Phonologique Internationale Tenue a Prague (18–21/xii 1930)* (pp. 79–96). Prague: Jednota Československých Matematiků a Fysiků.

Pouplier, M. (2012). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In S. Fuchs, M. Weirich, D. Pape & P. Perrier (Eds.), *Speech Planning and Dynamics* (pp. 63–96). Berlin: Peter Lang.

Pouplier, 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. DOI: https://doi.org/10.1016/j.wocn.2017.01.006

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org

Redford, M. (2008). Production constraints on learning novel onset phonotactics. *Cognition*, *107*, 785–816. DOI: https://doi.org/10.1016/j.cognition.2007.11.014

Ridouane, R. (2008). Syllables without vowels: Phonetic and phonological evidence from Tashlhiyt Berber. *Phonology*, *2*, 321–359. DOI: https://doi.org/10.1017/S0952675708001498

Rowley, A. (1990). North Bavarian. In C. Russ (Ed.), *The Dialects of Modern German* (pp. 417–437). London: Routledge.

Sancier, M., & Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436. DOI: https://doi.org/10.1006/jpho.1997.0051

Scholes, R. J. (1966). *Phonotactic Grammaticality*. The Hague: Mouton & Co. DOI: https://doi.org/10.1515/9783111352930

Seidl, A., Onishi, K., & Cristia, A. (2013). Talker variation aids young infants' phonotactic learning. *Language Learning and Development*, *10*(4), 297–307. DOI: https://doi.org/10.1080/15475441.2013.858575

Shockley, K., Sabadini, L., & Fowler, C. (2004). Imitation in shadowing words. *Perception and Psychophysics*, *66*(3), 422–429. DOI: https://doi.org/10.3758/BF03194890

Shosted, R. K., & Chikovani, V. (2006). Standard Georgian. *Journal of the International Phonetic Association*, *36*(2), 255–263. DOI: https://doi.org/10.1017/S0025100306002659

Smith, C. (1995). Prosodic patterns in the coordination of vowel and consonant gestures. In B. Connell & A. Arvaniti (Eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV* (pp. 205–222). Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511554315.015

Solé, M.-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech, 38*, 1–23. DOI: https://doi.org/10.1177/002383099503800101

Solé, M.-J. (1997). Timing patterns in the production of a foreign language. In L. Diaz & C. Pérez-Vidal (Eds.), *Views on the Acquisition and Use of a Second Language: EUROSLA'7 Proceedings* (pp. 539–551). Barcelona: Universidad Pompeu Fabra.

Steinlen, A. (2005). *A Cross-Linguistic Comparison of the Effects of Consonantal Contexts on Vowels Produced by Native and Non-Native Speakers*. Tübingen: Gunther Narr.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. DOI: https://doi.org/10.1016/j.wocn.2010.09.001

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics, 55*, 53–77. DOI: https://doi.org/10.1016/j.wocn.2015.11.005

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*(1), 47–62. DOI: https://doi.org/10.1177/002383099704000103

Weber, A., & Cutler, A. (2006). First-language phonotactics in second language listening. *Journal of the Acoustical Society of America, 119*(1), 597–607. DOI: https://doi.org/10.1121/1.2141003

Weinberger, S. (1994). Functional and phonetic constraints on second language phonology. In M. Yavaș (Ed.), *First and Second Language Phonology* (pp. 283–302). San Diego: Singular Publishing Group.

Whalen, D., Best, C., & Irwin, J. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics, 25*, 501–528. DOI: https://doi.org/10.1006/jpho.1997.0058

Wiese, R. (2000). *The Phonology of German*. Oxford: Oxford University Press.

Wiese, R., Orzechowska, P., Alday, P., & Ulbrich, C. (2017). Structural principles or frequency of use? An ERP experiment on the learnability of consonant clusters. *Frontiers in Psychology*, *7*(2005), 1–15. DOI: https://doi.org/10.3389/fpsyg.2016.02005

Wiesinger, P. (1989). *Die Flexionsmorphologie des Verbums im Bairischen*. Verlag der Österreichischen Akademie der Wissenschaften.

Wilson, C., & Davidson, L. (2013). Bayesian analysis of non-native cluster production. *Proceedings of NELS 40, MIT*, Cambridge.

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. DOI: https://doi.org/10.1016/j.jml.2014.08.001

Yanagawa, M. (2006). *Articulatory Timing in First and Second Language: A Cross-Linguistic Study* (Unpublished doctoral dissertation). Yale University.

Zampini, M. L. (2008). L2 speech production research. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 219–249). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/sibil.36.11zam

Zellou, G., Dahan, D., & Embick, D. (2017). Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience*, *32*(6), 776–791. DOI: https://doi.org/10.1080/23273798.2016.1275710

Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, *47*, 18–35. DOI: https://doi.org/10.1016/j.wocn.2014.09.002

Zsiga, E. C. (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, *28*, 69–102. DOI: 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*(3), 399–432. DOI: https://doi.org/10.1017/S0272263103000160

How to cite this article: Pouplier, M., Lentz, T. O., Chitoran, I., & Hoole, P. 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): 1, pp. 1–41. DOI: https://doi.org/10.5334/labphon.195

Submitted: 28 February 2019 **Accepted:** 22 November 2019 **Published:** 24 January 2020

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS