Journal of Phonetics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Special Issue: Mechanisms of regulation in speech, eds. Mücke, Hermes & Cho

Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness

Marianne Pouplier[a,*], Stefania Marin[a], Philip Hoole[a], Alexei Kochetov[b]

[a] Institute of Phonetics and Speech Processing, Schellingstr. 3, LMU, 80799 Munich, Germany
[b] Department of Linguistics, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada

ARTICLE INFO

Article history:
Received 20 April 2016
Received in revised form 17 January 2017
Accepted 19 January 2017
Available online xxxx

Keywords:
Russian onset clusters
Speech rate
Auditory cue robustness
Phonotactic frequency
Typological preference
Coarticulation
Durational control

ABSTRACT

This study presents data on the durational and timing characteristics of Russian onset clusters and their change as a function of speaking rate. The focus is on Russian due to it being known to have relatively less consonant overlap compared to languages like English and due to its unusual range of consonant clusters. Using articulography, we investigate whether these characteristics have implications for the flexibility of clusters under speech rate changes. In particular we ask whether a cluster's signal modulation profile, taken as an index of auditory recoverability, predicts the degree to which the overlap pattern of a cluster changes with rate. Previous research suggests that stop + stop clusters may be less susceptible to rate change than other, auditorily more robust clusters within the same language. Moreover, even though frequency and phonotactic preference are usually closely aligned, Russian also allows us to probe frequency effects on cluster timing, since for our data these factors are dissociated to a certain degree. Results show that both duration and relative timing of the consonants in a cluster are subject to change. Speech rate effects do not scale uniformly throughout the cluster but are carried predominantly by the constriction formation duration of C2. Clusters show a decreasing rate effect from high to low frequency clusters. Grouping clusters according to their assumed perceptual robustness does not lead to a clear result. We discuss these findings in the context of models of durational control of speech production.

1. Introduction

The goal of the current paper is to shed light on how timing and overlap in consonant clusters are affected by speech rate. More specifically, we ask whether either frequency or auditory cue robustness may interact with the relative flexibility clusters exhibit under rate changes in conjunction with language-specific phonotactic properties. To that effect, we present articulatory data on the durational characteristics and coarticulatory patterns of Russian onset clusters in two speaking rate conditions. The study addresses three overall research questions: For one, we investigate the extent to which onset clusters in Russian, a language with relatively little consonantal overlap, adapt to speech rate variation. This stands in the context of our increasing knowledge of systematic cross-linguistic differences in coarticulation and consonant-consonant timing. While languages like English, German, or French feature a relatively high degree of overlap between sequences of consonants, other languages like Russian have been reported to time their consonants comparatively far apart, which may in certain circumstances give rise to a transitional schwa between consonants (Bombien & Hoole, 2013; Davidson & Roon, 2008; Zsiga, 2000, 2003; relatedly Öhman, 1966). As explained in more detail below, there is reason to expect that clusters in a language with a low overlap pattern such as Russian may exhibit resistance to rate changes and this is one general hypothesis we put to test.

Secondly, we pursue the possibility that a cluster's susceptibility to rate changes is related to its cue robustness. In particular, we follow an idea originally proposed by Wright (1996) that clusters which are problematic in terms of auditory recoverability fail to increase overlap with rate. Russian offers an ideal opportunity to probe this possibility further since from a typological viewpoint, Russian onset cluster phonotactics are unusual. Russian not only features the cross-linguistically most common sonority-rising clusters like /bl-/, but also sonority pla-

---

* Corresponding author.
E-mail addresses: pouplier@phonetik.uni-muenchen.de (M. Pouplier), marin@phonetik.uni-muenchen.de (S. Marin), hoole@phonetik.uni-muenchen.de (P. Hoole), al.kochetov@utoronto.ca (A. Kochetov).

http://dx.doi.org/10.1016/j.wocn.2017.01.006

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

teau (/tk-/) and even sonority falling sonorant-obstruent clusters like /lb-/. Henke, Kaisse, and Wright (2012) have argued that the traditional concept of sonority, designed to capture cross-linguistic segment-sequencing preferences (e.g., Clements, 1990), is epiphenomenal to auditory cue robustness. The current work empirically tests their prediction that clusters with low cue robustness should resist speech rate changes. Lastly, we ask whether differential cluster sensitivity to speech rate variation may be related to cluster frequency. While Russian does not permit us to completely de-correlate frequency and cue-robustness, the grouping of the clusters we have recorded is not isomorphic for these two factors and we can thus begin to assess their relative contribution to articulatory dynamics.

The following sections will consider in turn the state of the art for how speech rate changes are manifest in articulation in conjunction with known cross-linguistic coarticulatory differences, proposals on how rate may interact with auditory cue robustness, and finally current knowledge on the effects of phonotactic frequency.

1.1. Speech rate changes and the language-specific nature of coarticulation

A series of papers has investigated speech rate with the particular goal of understanding the fundamentals of durational control in speech production and aspects of movement optimization (e.g., Adams, Weismer, & Kent, 1993; Nelson, Perkell, & Westbury, 1984; Ostry & Munhall, 1985; Shaiman, Saltzman, & Tuller, 1995; Smith, Goffman, Zelaznik, Ying, & McGillem, 1995; Tasko & Westbury, 2004). All of these studies have revealed complex trading relations between velocity and amplitude, either of which may or may not vary under rate scaling (see Berry (2011) for an overview). The literature on speech rate effects has mainly been concerned with consonant–vowel transitions and rate effects on individual constrictions, in particular vowels (e.g., Lindblom, 1963). Speech rate scaling is not uniform throughout the CVC syllable in that the proportionally greatest durational compression occurs in the nucleus, and the least in the onset consonant (Gay, 1978, 1981). Adams et al. (1993) found that even the movement phases within a segment differ in how consistently rate changes occur: release gestures exhibited more uniform rate scaling compared to constriction formation gestures of the same segment. They attributed this to the CV transition being perceptually more important; therefore in their view control strategies may seek to minimize variation for perceptually critical movement phases (“islands of reliability”, Adams et al. (1993, p. 48)). Yet they acknowledge that stress may have been a confounding factor in their results. Another known effect of faster speech rate is increased coarticulation in terms of earlier anticipatory movement onsets between successive segments (Agwuele, Sussman, & Lindblom, 2008; Gay, 1978, 1981). Relatedly, inter-articulator phasing has been shown to change with rate (Nittrouer, 1991; Nittrouer, Munhall, Kelso, & Tuller, 1988).

Generally, there has been very little work on how speaking rate affects the timing of consonant sequences, which is the main focus of our present work. In an EPG study, Byrd and Tan (1996) found for American English cross-word consonant clusters that speaking rate changes have a linear effect on the duration of each member of the C1#C2 sequence. In contrast to this, rate conditioned changes to overlap were not consistently present in the data which lead the authors to suggest that listener-oriented factors may act as a constraint on overlap. The interaction of listener-oriented factors with rate variation was targeted specifically in a study on clusters in Tsou (Wright, 1996). Based on acoustic measures Wright observed a greater degree of rate-conditioned shortening for C1 compared to C2, but due to his experimental design (acoustic recordings of words in isolation) the duration measures could not be applied to stop-initial clusters. Hence his study could not distinguish whether this result was due a manner effect (C1 were non-stops and C2 was always a stop in his stimuli for this part of his analyses) or a position effect or an interaction between the two. For stop + stop clusters, C1 retained a release burst at all speaking rates with little variation between rates in the inter-burst-interval (temporal interval between the two stop bursts). At the same time, Wright could trace some acoustic influence of C2 on the release burst spectrum of C1, providing evidence for articulatory overlap between the consonants, despite a consistently released C1 even at the fast rate. Wright concluded that consonant cluster overlap generally is constrained by perceptual factors, since in his data only clusters with particularly vulnerable auditory cues (stop + stop sequences) barred an increase in overlap in fast speech.

Expanding on Wright (1996), Henke et al. (2012) proposed that gestural timing is a key factor in understanding how languages may achieve some level of cue robustness for clusters with perceptually sub-optimal signal modulation profile (discussed in detail in the next section). By extension, these clusters are predicted to be rather immune to any coarticulation-scaling factors such as speech rate. Their proposal thus concerns cluster-specific variation to overlap patterns within a given language: the existence of typologically exceptional (sub-optimal cue) clusters within a language is seen as conditional on these clusters showing a different coarticulation pattern. There are several proposals that go further in suggesting that languages featuring unusually complex consonant phonotactics are generally characterized by certain coarticulatory patterns. That coarticulation is language specific has been known at least since Öhman’s influential (1966) paper in which he observed a relative weakness of V-to-V coarticulation in Russian compared to Swedish and English. More recent work has accumulated more and more evidence for the language-specific nature of both V-to-V and C-C coarticulation, which we here take to be tantamount to temporal overlap (Beddor, Harnsberger, & Lindemann, 2002; Bombien & Hoole, 2013; Chitoran, Goldstein, & Byrd, 2002; Flemming, 2011; Kochetov, Pouplier, & Son, 2007; Ma, Perrier, & Dang, 2015; Manuel, 1990; Marin & Pouplier, 2014; Mok, 2010; Pouplier & Beňuš, 2011; Zsiga, 2000, 2003). In addition, there have been several proposals that there is a causal relationship between certain phonological phenomena and the coarticulation patterns within a language. For instance, Ma and colleagues (2015) recently attributed language specific effects in vowel-to-vowel coarticulation in Mandarin and French to the role of the syllable as a prosodic domain. In contrast to French, V-to-V coarticulation is effectively blocked in Mandarin, consis-

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

tent with the view that the syllable is an independent planning domain in Mandarin but not in French. Since planning domains in motor control are the domains in which forthcoming movement is planned, coarticulation will be more extensive within rather than across planning units. Relatedly, Pouplier and Beňuš (2011) argued in the context of Slovak syllabic consonants that the range of consonant clusters permitted in a language is related to the language-specific degree of overlap between consonants, with more cluster types being attested in languages where consonants in a cluster are timed further apart, i.e. with greater consonant-consonant lags (see also Chitoran, 2016). In a similar vein, Bombien and Hoole (2013) proposed that languages' voicing specification (e.g. whether stops are fully voiced or not) may interact with preferred overlap patterns. Evidence for the view that there is a principled relationship between phonotactics and consonant timing comes from certain acoustic characteristics of languages such as Georgian, Russian, Tsou and others which feature a typologically unusual range of clusters: These languages are not only known for systematically released stops but also for the presence of transitional schwas between consonant sequences, even in those with an unproblematic auditory profile. Generally, there is thus some indication that 'permissive' cluster phonotactics co-occur with a particular coarticulation pattern of relatively less consonant-consonant overlap not only in the particular typologically unusual clusters themselves (such as stop + stop clusters), but rather the low overlap patterns may characterize these languages as a whole. That is, not only do clusters differ within a language in their relative overlap, but languages differ in what one might term a basic coarticulatory setting. It may then be the case that a relative immunity of clusters to rate variation may permeate the language as a whole as part of this hypothesized coarticulatory setting, and the results of our study will be informative about this question. There is some limited previous work on prosodic variation of durational patterns in Russian which might be taken as circumstantial evidence in this direction. Consonantal duration patterns for Russian have previously been investigated acoustically in interaction with word-edge structure, testing the extent to which duration serves as a listener-oriented cue to boundary location. Redford (2007) found that prevocalic stops are generally longer than postvocalic ones, yet in Russian word boundary effects on duration were less pronounced compared to English. Davidson and Roon (2008) observed for Russian that C1 duration for stops remained constant under changes in syllable structure (C1#C2, #C1C2, #C1əC2); they conjectured that this was due to stops being mostly audibly released in Russian (Zsiga, 2003). That Russian clusters may not be immune to rate variation is, however, suggested by a small articulatory study on rate effects on Russian C1#C2 cross-word stop + stop clusters by Kochetov et al. (2007) with data from only two speakers. This work reported a consistent shortening of both consonants at fast rate, but the results for plateau lag between the consonants were inconsistent in that one of the two recorded speakers showed a rate effect but the other one didn't. They again mention the requirement for stops to be released in Russian as a possible factor constraining rate variation. Our study will further contribute to this discussion, and we will now turn in more detail to proposals for a perception-based taxonomy of clusters.

1.2. Auditory cue robustness of clusters

One of the goals of our paper is to investigate a possible interaction between auditory cue robustness and speech rate. The literature on the characteristics of speech rate changes has, as outlined in the preceding section, repeatedly appealed to perceptual recoverability as a constraining factor on articulatory dynamics whenever non-uniformities in durational effects have been observed. Likewise it has been argued that the existence of certain clusters, tied to their perceptual recoverability, is conditional on a particular (relatively lower) overlap pattern. By extension of this argument we can expect a relative inflexibility of these clusters to rate changes. This section discusses previous proposals on which criteria may best be employed to judge the perceptual recoverability of a cluster. The working definition of cue robustness applied in the current study builds on several proposals which all share a core insight: the traditional concept of sonority is seen as epiphenomenal to the preference for sequences of segments to follow an auditorily optimal signal modulation profile (Chitoran, 2016; Henke et al., 2012; Mattingly, 1981; Ohala & Kawasaki-Fukumori, 1997; Wright, 1996, 2004). In the following, we detail the relevant proposal of Henke et al. (2012) since they make concrete predictions about how particular clusters should fare with respect to perceptual cue robustness and hence susceptibility to temporal variation in overlap. Instead of assuming a context-free segmental sonority hierarchy based on manner (e.g., vowel > glide > liquid > nasal > obstruent (Clements, 1990), or some more detailed version thereof), the authors maintain that the cross-linguistic (dis-)preferences for certain phonotactic patterns can be more adequately captured on perceptual robustness considerations. Thereby key factors are the internal cue strength of segments as well as the extent to which cues to a given segment's identity are carried or obscured by transition dynamics to neighboring segments. Cue strength is defined in terms of signal intensity, temporal spread of cues, and the degree of amplitude and spectral modulation in a sequence of segments. High signal intensity, a large temporal spread of cues, and a high degree of amplitude and spectral modulation guarantee cue robustness. Cue robustness is therefore contextually specific, since what may be robust in one context may not be so in another. For example, sibilants, on this account, have extremely salient internal cues to both place and manner, making them preferred candidates for syllable-edge positions. However, a sequence of two sibilants will have poor amplitude and spectral modulation, making sequences such as /sʃ/ fare worse auditorily than, say, /sp/. Stops preferably precede a vowel since they have poor internal cues to place, making /sp/ fare better than /ps/. Citing as examples Tsou (Wright, 1996), Georgian (Chitoran et al., 2002) and Montana Salish (Flemming, Ladefoged, & Thomason, 2008), Henke and colleagues substantiate their argument that the existence of perceptually difficult sequences in a language entails a limit on gestural overlap for these clusters. The remedy for poor cue robustness is to limit coarticulation with neighboring sounds.

Wright's (1996) dissertation on consonant clusters in Tsou pioneered the idea that speech rate manipulation can be used to corroborate the hypothesis that auditory cue preservation constrains consonantal overlap. As a corollary, these clusters

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

should exhibit little timing flexibility under rate changes (since fast speech is otherwise known to cause increased coarticulation). His acoustic measurements provided some evidence that stop + stop clusters fail to increase overlap at faster speech rate while clusters with C1 fricative, affricate, or nasal in the same language do. But given the methodological limitations (discussed in the previous section), the acoustic measures he could use are only indirectly informative about consonantal overlap and rate flexibility. Nonetheless his work provides some evidence for the hypothesis that auditory cue preservation may interact with speech rate. Also our own previous work which reported preliminary results for a subset of the speakers presented in the current paper gave evidence for an interaction between cluster type and rate flexibility (Pouplier, Marin, & Kochetov, 2015). That study partitioned clusters into rising (e.g., /bl/), plateau (e.g., /mn/), or falling (e.g., /lb/) sonority profile groups following a traditional manner-based sonority scale. There was a significant interaction between sonority group and speech rate caused by a higher degree of overlap at the fast rate for the rising and plateau, but not for the falling sonority group. We concluded that sonority was, however, not the best main predictor for the results: If non-canonical sonority had been the key factor, the plateau and falling clusters should have patterned together and contrasted with the rising/canonical profile group. In the current study, we take up Henke et al.'s argument that sequencing rules for consonants are better expressed in terms of cue robustness and test whether we find a more consistent result.

One can hardly talk about phonotactic preference without mentioning frequency. These two factors stand in a strong implicational relationship to each other, making arguments that one has precedence over the other usually circular. And while we do not claim that we can propose a solution to this long-standing problem here, the frequency distribution of our selection of Russian clusters does not align 1:1 with their cue robustness, and we can therefore gain some understanding of how these factors may relate to each other. We will now briefly review known effects of frequency on speech production before turning to our actual experiment.

1.3. Frequency effects

Frequency effects in speech production have widely been demonstrated at several processing levels and in terms of different parameters such as word and syllable frequency, phonotactic probability, and collocational probability (among others, Aichert & Ziegler, 2004; Aylett & Turk, 2006; Edwards, Beckman, & Munson, 2004; Gahl, 2008; Goldrick & Larson, 2008; Jaeger & Hoole, 2011; Jurafsky, Bell, Gregory, & Raymond, 2001; Lin, Beddor, & Coetzee, 2014; Munson, 2001; Munson & Solomon, 2004; Stephenson, 2003; Vitevitch, Armbrüster, & Chu, 2004). It is well known that more frequent words tend to have reduced acoustic durations, less extreme articulator positions for consonants, centralized vowels, and also fare better on accuracy, speed of production, and fluency metrics. Goldrick and Larson (2008) provided some limited evidence that frequency affects production independently of inherent phonetic complexity since their experiment could manipulate speech error patterns based on an implicit learning paradigm. Error patterns for the same segment varied as a function of the phonotactic probability distribution that experimental participants were exposed to due during an implicit learning phase. The authors argue that no such frequency-based pliability of patterns was to be expected if a segment's phonetic complexity conditioned error patterns. Despite the large body of work on frequency effects, how exactly frequency may affect articulatory dynamics other than articulator position is by and large unknown. Tomaschek and colleagues (Tomaschek, Wieling, Arnold, & Baayen, 2013) investigated articulatory changes to German vowels as a function of word frequency and found some indication that higher word frequency conditioned an earlier movement onset of the final coda consonant during the vowel. Relatedly, it has been demonstrated that words in high density lexical neighborhoods are produced with increased coarticulation (Scarborough, 2012). Overall this means that we can expect high frequency clusters to be produced with more overlap. Whether phonotactic frequency interacts with speech rate with respect to coarticulatory dynamics has, to our knowledge, not been investigated so far and will be one of the contributions of the current study.

In sum, the goal of our study is to use Russian onset clusters to study the effect of speech rate on consonant-consonant timing, both globally and specifically for particular cluster groups. First, we group clusters according to their auditory cue robustness and seek to replicate the findings of Wright (1996) that in perceptually sub-optimal clusters consonant overlap fails to increase under rate scaling. Secondly we ask whether cluster frequency conditions differential effects on overlap changes under rate.

2. Method and materials

2.1. Data acquisition and treatment

We recorded articulography (EMA, Carstens AG501) data with a sampling rate of 1.25 kHz together with synchronized audio at 25.6 kHz. Standard calibration and position recovery methods were used. Receiver coils were attached to the tongue (tip, mid, back), jaw, upper and lower lips, plus four reference sensors allowing for the correction of head movement and rotation to the occlusal plane. As part of position recovery, the kinematic data of the tongue tip coil were filtered at 40 Hz, the reference coils at 5 Hz, and all others at 20 Hz. For each speaker, we corrected for head motion, rotated the data to the occlusal plane, and obtained an outline of the hard palate.

2.2. Speakers

12 native speakers of Russian participated in the study. All of them lived in Munich, Germany, at the time of the recording but reported speaking Russian on a regular basis. A native speaker assistant was present during the recordings and conducted all interactions with the participants in Russian. None of the participants reported any speech or hearing deficits.

2.3. Stimuli

Twelve different #C1C2V syllables were recorded with V=/a/ except for V = /o/ for /mn-, kt-/ due to phonotactic constraints.

Thus all stimuli were possible syllables of Russian. Clusters were originally selected such that we could construct equal-size groups for falling, rising, or plateau sonority profile groups (Pouplier et al., 2015).[1] As laid out in the Introduction, we focus in this paper on a cue-based approach to cluster phonotactics and deduce our grouping of clusters from Henke et al.'s (2012) proposal on contextual cue robustness, specifically Section 2 of their article. Therefore our groups are not perfectly balanced. The operational grouping of stimuli is given in Table 1. /ʃp, ʃm/ are considered optimal because, as mentioned before, sibilants are ideal candidates for syllable-edge position due to their strong internal cues. Stops and nasals preferably occur close to the vowel; in addition both clusters have a clear signal modulation profile. /bl, gl/ ensure optimal cue recovery for the lateral since laterals depend to a large degree on the formant transitions into the vowel while laterals themselves are able to carry place information for the stops. Given that laterals depend on transitional information for full recovery, /lb, lg/ are grouped as medial: stops are not able to carry transitional information of the laterals. At the same time, the stops are in optimal pre-vocalic position. The lateral is in optimal position for /ml/, but nasals have very poor internal place cues (worse than stops in this respect) and are in a less-than-ideal place if not next to a vowel. Also signal modulation is attenuated in this cluster. Signal modulation is even worse in stop + stop sequences and moreover, stops are highly dependent on formant transitions for place recovery. Therefore /tk, kt/ are clearly clusters with low auditory cue robustness. /mn/ is similarly problematic since nasals are neither strong on internal place cues nor good carriers of a neighboring segment's cue. In addition, the signal modulation profile is flat. Finally /mx/ has the nasal in a poor recovery position for the named reasons; the same holds for the fricative in /xm/ since it is in a poor recovery position before the nasal. Due to their low signal modulation profile we also place these latter two clusters in the poor cue robustness group.

For the frequency analyses, cluster frequency was calculated on the basis of the Russian Internet Corpus available from the Centre for Translation Studies at the University of Leeds (Sharoff, 2005, 2006); this is a corpus of about 90 million words (198,509,029 tokens or 2,771,231 types/lemmas). The 100,000 most frequent words are available for the corpus with the lowest listed frequency being 98ipm. We used a log normalized frequency of the summed occurrences of a given cluster in absolute word-initial position to partition the clusters into three frequency groups. The frequency range of our clusters is 6–12.7. In order to ensure comparability to the auditory cue robustness analyses, we operationally binned the clusters into three frequency groups, ‘high,’ ‘mid,’ and ‘low’ such that they render three groups of 4 clusters. Table 2 gives the stimuli binned into these groups with each cluster's frequency and average group frequencies in parentheses. Since frequency is a quasi-continuous variable in our stimuli, we also test, for the key findings, whether our results hold when entering frequency as a continuous independent variable.

Table 1
Stimuli by Auditory Cue Robustness Group and number of tokens that entered into the analysis for each cluster. All stimuli are legal syllables of Russian.[1]

Optimal cue robustness (O-Group; token total = 426)	Medium cue robustness (M-Group; token total = 322)	Poor cue robustness (P-Group; token total = 412)
bla (n = 106)	lba (n = 107)	tka (n = 97)
gla (n = 100)	lga (n = 106)	kto (n = 95)
ʃpa (n = 110)	mla (n = 109)	mxa (n = 43)
ʃma (n = 110)		xma (n = 69)
		mno (n = 108)

[1] Sample Russian words with these clusters include: *blaga* ‘good (noun)’, *glatkij* ‘smooth’, *ʃpaga* ‘sword’, *ʃmara* ‘hooker’, *lba* ‘forehead (gen.sg.)’, *lgal* ‘(he) lied’, *mlatʃɨj* ‘junior’, *tkal* ‘wove’, *kto* ‘who’, *mxa* ‘moss (gen. sg.)’, *xmarʲ* ‘haze’, *mnoga* ‘many.’ Here and throughout the paper Russian forms are transcribed broadly and reflect neutralizing changes such as unstressed vowel neutralization, voicing assimilation, and palatalization.

Table 2
Stimuli binned into three frequency groups. The summed log frequency for each cluster is given in each cell. The average frequency for each group is given in the header line.

High (12.2)	Mid (8.8)	Low (6.7)
gl (12.7)	ml (9.5)	ʃp (7.2)
mn (12.4)	tk (8.9)	lg (7.0)
kt (12.1)	lb (8.8)	ʃm (6.5)
bl (11.5)	xm (7.9)	mx (6.0)

2.4. Rate elicitation

The target syllables were embedded in the carrier phrase: [ˈgromka ____ paftarʲil], ‘(He) repeated ___ loudly’. Two speaking rates were elicited: Speakers saw a horizontal bar moving across the screen that indicated the time window within which they had to say the entire utterance. Bar duration was set to 2.3 and 1.5 s for the slow and fast rate respectively. During the practice phase of the experiment each speaker was offered the possibility to adjust the rates to a range suitable for them. None of the speakers requested a change in bar duration for either rate. Speakers then practiced the two rates before the actual recording began. Five repetitions per rate were recorded; the rates were blocked and speakers were alerted to rate changes. The first stimulus of each rate block was a dummy and served to remind speakers of the targeted rate. Each repetition of the two rate blocks was preceded by a block of stimuli for a different experiment; whether the slow or fast stimulus block came first was varied for each cycle through the stimuli.

For one speaker, the fast rate could not be analyzed since she reduced the final syllable of the carrier phrase word *gromka* to such an extent that neither labial nor velar-initial clusters could be reliably identified. Since our research question relies on contrasting the two speaking rates, the data from this speaker were excluded as a whole for present purposes. The targeted token total for the remaining speakers amounted to 11 speakers × 12 clusters × 5 repetitions × 2 rates = 1320. Data loss occurred for 160 tokens, leaving 1160 tokens for analysis (578 fast, 582 slow rate). This happened for the following reasons: For several of the speakers, entire clusters had to be removed from analysis because articulatory segmentation could not be performed adequately. This in turn was due to a number of our speakers having a quite posterior, rather uvular articulation for dorsals (/g, k, x/). /g/ was prone to a uvular articulation especially in the context of /l/, which is a

---

[1] The data of the Pouplier et al. (2015) publication are a subset of the data presented here since at the time point of the earlier publication we had recorded five out of our targeted 12 speakers. The key result of the 2015 publication generalizes to the larger dataset reported here: There is an interaction between rate and sonority profile due to the rising and plateau, but not the falling profile group showing increased consonant overlap with rate.

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

dark liquid in Russian. During data segmentation it then became apparent that the sensor glued to capture velar articulations was positioned for some of these speakers too anterior to capture their velar (uvular) constrictions in all conditions and some of the clusters containing velars had to be excluded from analysis. Partial data exclusion for this reason occurred for 5 of our 11 speakers. The fricative /x/ was affected to a greater degree than /k, g/. Incidental data loss caused by speech errors or technical recording failure occurred for 9 tokens. Table 1 gives the number of tokens included in the analysis for each cluster.

2.5. Measurements and statistics

For each sensor time series of interest, the articulatory constriction formation and release was identified based on the first derivative of the position signals. The velocity signals were smoothed at 24 Hz with a moving average filter. For each gesture, first the time points of the peak velocities of the constriction formation and release movements were identified algorithmically. Then a 20% threshold of the peak velocity was used to determine the time points of movement onset, achievement of target, and release. Movement onset and target achievement were determined for each consonant by a 20% threshold of the peak velocity of the constriction formation; the release time point was computed relative to the peak velocity of the constriction release (see Fig. 1a). To measure labial constrictions, lip aperture was computed as the Euclidean distance between the upper lip and lower lip time series. 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.

The time point of movement onset can in some cases be difficult to determine, particularly for C1 in a C1C2 cluster. This occurs mainly when there is a rather flat, prolonged rise towards the peak velocity. In these cases, the movement onset is identified unreasonably early by the algorithm and this measurement time point was excluded from analysis ($n = 63$, of which $n = 56$ for C1). Another known segmentation difficulty arises with the presence of multiple velocity peaks. In those cases, we forced the algorithm to pick the largest velocity peak.

Articulatorily, we employed several measures to assess speaking rate effects on each consonant's duration and overlap.

(1) Durational measures evaluated as raw values and relative to cluster duration (cf. Fig. 1a):
  - *Constriction formation duration* was defined as the interval from onset of movement for the constriction formation to the target achievement. Tokens for which movement onset could not be determined (see above) do not factor into this analysis.
  - *Plateau duration of each consonant* was defined as the interval between target achievement and release.

To assess proportional duration for these two measures, we normalize for cluster duration defined as the interval between the peak velocity of constriction formation for C1 to target release of C2. Peak velocity of C1 was chosen for normalization rather than movement onset to prevent unnecessary data loss. We can thus determine normalized plateau duration also for clusters for which we do not have a reliable movement onset of C1. The measurements are illustrated in Fig. 1b.

(2) Relative timing measures evaluating the timing of C2 relative to C1's movement cycle (Fig. 1c).

Relative timing measures between the two consonants were computed by setting all segmentation landmarks of C2 relative to the [0, 100] duration of C1, the latter being defined as the time interval between peak velocity of the constriction formation and release of that consonant. Again peak velocity of C1 was chosen for normalization rather than movement onset to prevent unnecessary data loss. We can thus include in our analyses relative timing measures for clusters for which we do not have a reliable movement onset of C1 but for which all other segmentation time points are available.

- *Plateau lag* was defined in absolute time as the interval between release of C1 and target achievement of C2. In normalized time, it was defined as the normalized target achievement of C2. Since C1 release was taken as the end-point of our normalization interval, values greater than 100 mean that C2 target achievement occurs after the articulatory C1 release.
  For plateau lag we report both absolute and normalized values.
- *Onset lag* captures when C2 begins its movement within C1's normalized duration. This is conceptually equivalent to computing a phase value of C2 onset relative to C1.

The proportional measures are illustrated schematically in Fig. 1c for three different relative timing scenarios.

Each trial was also segmented acoustically in order to verify that speakers actually distinguished the speech rate conditions at a global level. The duration of the entire utterance was defined as ranging from the burst of the initial velar (*gromka*) to the end of periodic vibration for the final liquid (*paftarʲil*). Target stimulus duration was defined as the interval from end of the word-final vowel in *gromka* to the beginning of closure for the initial labial in *paftarʲil*.

For statistics we used linear mixed models in R (R Core Team, 2015) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). All models were specified for random intercept for Speaker and for Cluster as well as by-Speaker random slopes for Rate. P-values were obtained by a likelihood-ratio-test in which the full model was compared to a model without the factor in question. Tukey posthocs were conducted with the multcomp package (Hothorn, Bretz, & Westfall, 2008), effects were extracted using the effects package (Fox, 2003).

3. Results

3.1. Global rate effects

The actual speaking rate for each trial was determined based on a syllables per second calculation. We measured for each trial from the acoustic signal the duration of the entire sentence and calculated speaking rate as (nominal) syllables per second. The across-subject syllables per second mean was 5.4 (SD = 0.6) for the fast and 4.4 (SD = 0.7) for the slow

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

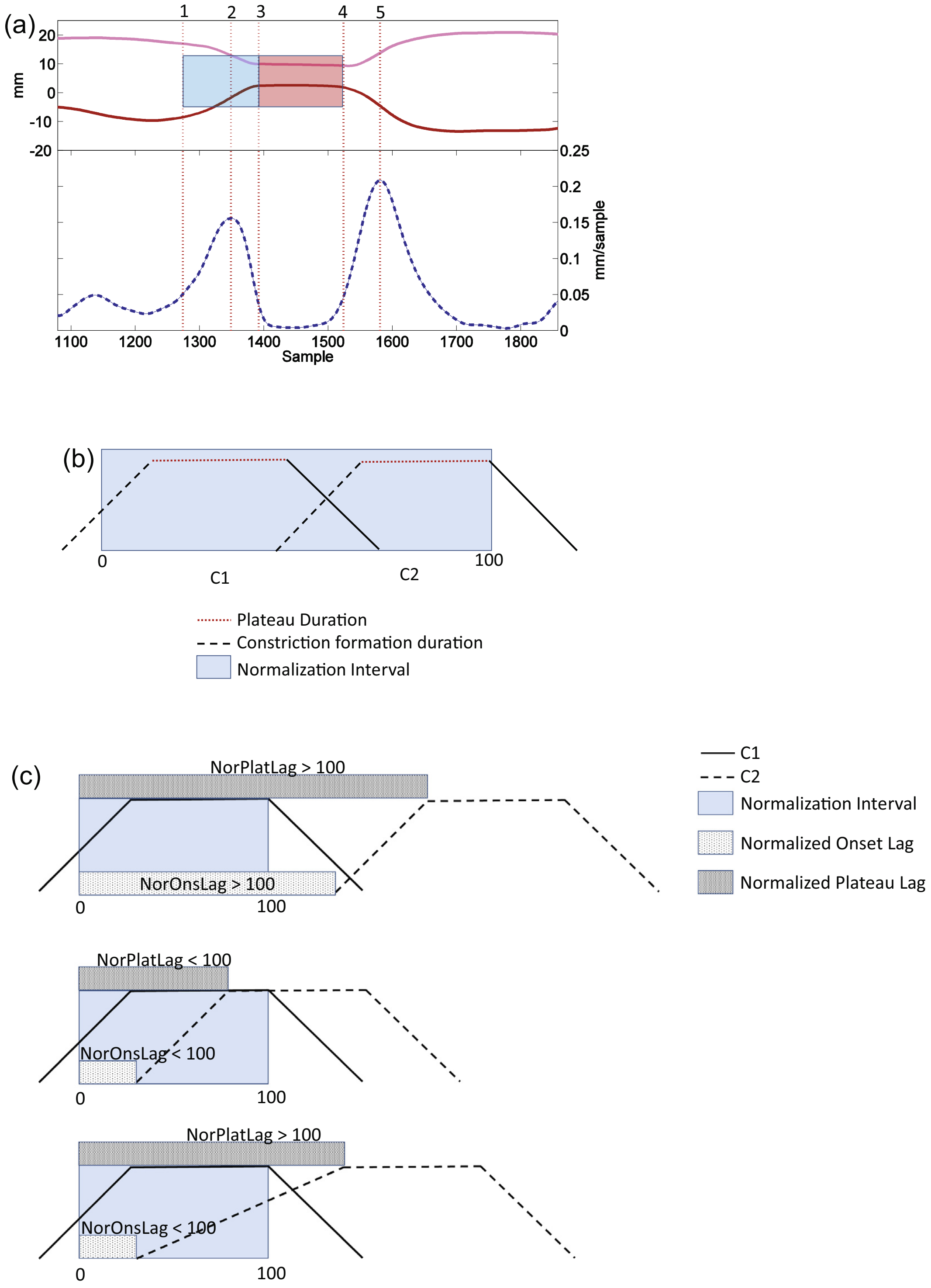

Fig. 1. (a) Illustration of segmentation conventions. The upper panel shows the time series of tongue tip position (top line: horizontal, lower line: vertical), the lower panel shows the corresponding tangential velocity. The dotted vertical lines indicate the segmentation time points (1) movement onset, (2) peak velocity of constriction formation, (3) target achievement, (4) release, (5) peak velocity of release. The blue/lighter shaded box in the position time series is the constriction formation duration, the red/darker box indicates the plateau duration of this example constriction. (b) Schematic illustration of durational measures and the normalization interval cluster duration (peak velocity of constriction formation C1 to release of C2). (c) Schematic illustration of relative timing measures for three different overlap scenarios. The plain shaded box shows the time interval of C1 relative to which timing was proportionally evaluated (from peak velocity to release of C1). In the top panel, there is no overlap between the two consonants, both normalized plateau lag and onset lag are >100. The middle panel illustrates a high degree of overlap with both measures being <100. The bottom panel illustrates a case in which there is a high degree of overlap at movement onset (Normalized Onset Lag < 100), but no overlap of the plateaus (Normalized Plateau Lag > 100) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

condition. A linear mixed model on the dependent variable syllables per second was run with fixed factors Rate and Repetition, with random effects as specified above. Rate ($\chi^2(1) = 17.67$, $p = 0.00003$) and Repetition ($\chi^2(1) = 88.02$, $p < 0.00001$) were significant, the interaction was not ($\chi^2(1) = 1.18$, $p = 0.28$). The estimated effect of Rate is a decrease

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

of 0.97 (se ±0.15) syllables/s at the slow rate, the estimated effect of Repetition is an increase of .05 (se ±0.006) syllables/s with repetition. We further tested for local speech rate differences between conditions by running the same statistical model on acoustic target stimulus duration. The mean for the slow rate was 0.32 s (SD 0.073) and 0.24 s for the fast rate (SD 0.038). Again there was a significant effect of Rate ($\chi^2(1) = 12.82$, $p = 0.00034$), but not of Repetition ($\chi^2(1) = < 1$, $p = 0.64$). The Rate*Repetition interaction was significant ($\chi^2(1) = 6.3$, $p = 0.012$). The estimated effect of going from fast to slow rate is an increase of 80 ms (se ±16 ms). The significant interaction arose due to a small increase in durational difference between the two speaking rates over the course of the experiment (Fig. 2) which in turn is due to duration decreasing as a function of repetition in the fast but not in the slow rate. Crucially, Tukey posthocs confirmed a statistically significant slow-fast difference between all paired repetitions (fast.1 – slow.1, fast.2 – slow.2, etc.). Overall, our experimental setup therefore succeeded in eliciting two rate conditions both at the phrase and at the target word level and rate condition differences are present throughout the experiment regardless of repetition. We therefore do not consider the factor Repetition any further in the following analyses. The remaining parts of our paper will focus on the effects of speech rate on consonant-consonant overlap in interaction with auditory cue profile and/or frequency.

3.2. Global effects: durational characteristics of C1 and C2 as a function of speech rate

In this section, effects of speaking rate on the movement cycle of the individual consonants are considered. This serves to gain a global understanding of how rate effects are manifest in clusters before considering possible interactions with cluster type in the remainder of the paper. For each individual cluster member we look at the duration of constriction formation (CLO), plateau duration (PLAT) and total duration (CLO + PLAT). Table 3 gives the averages and standard deviations. As can be seen from the table, both consonants shorten overall with increasing rate and both constriction formation and plateau duration are shortened. We note very similar total durations of C1 and C2, yet C2 has a slightly shorter plateau compared to C1 but a longer constriction formation duration. Three mixed models were run, one for each measure as dependent variable. Independent variables were Position (C1, C2) and Rate (slow, fast). The interaction reached significance for neither closure formation nor plateau duration nor total duration (for all three: $\chi^2(1) < 1$). Comparing a Rate + Position model against a model with only Position underscored the significant Rate effect for all measures (closure formation duration: $\chi^2(1) = 12.31$, $p = 0.0004$, plateau duration: $\chi^2(1) = 12.4$, $p = 0.0004$, total duration: $\chi^2(1) = 13.2$, $p = 0.0003$). Position (C1, C2) was significant for plateau duration ($\chi^2(1) = 74.32$, $p < .0001$), closure formation duration ($\chi^2(1) = 68.92$, $p < 0.0001$) alike and approached significance for total duration ($\chi^2(1) = 3.43$, $p = 0.06$). This confirms our qualitative observations statistically: constriction formation duration for C2 is longer than constriction formation duration for C1, but vice versa for plateau duration. The durational position effect is independent of rate (lack of an interaction).

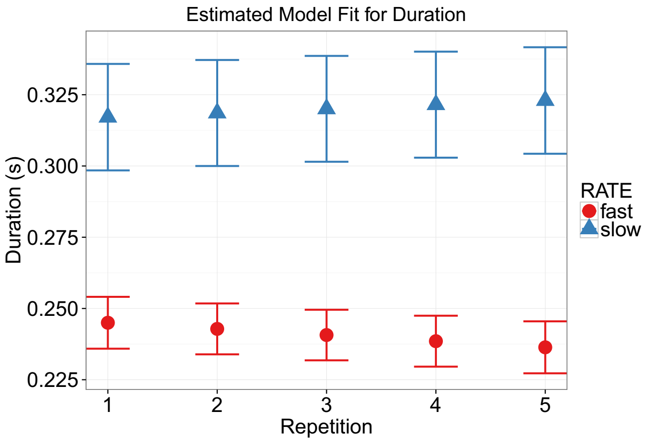

Fig. 2. Estimated effects for the Repetition-Rate interaction for acoustic target stimulus duration. Whiskers show the 95% confidence interval. Results show a clear rate separation for all repetitions.

At first blush, the position effect may be attributed to intrinsic consonant duration differences, but recall that our corpus contains obstruents as well as sonorants in either position and thus the usual confound between position and segmental identity is alleviated to a considerable extent in our present dataset. When looking at the correlation between closure formation duration and plateau duration for C1 compared to C2, we again find a systematic difference as a function of position. Fig. 3 presents a scatter plot for C1 and C2 each, across all clusters and subjects, plotting plateau duration as a function of closure formation duration. While there is a (modest) positive correlation for C1 for both rates, there is no correlation for C2 between these two measures at either rate. At the fast condition the pairwise Pearson correlation coefficients are for C1 $r = 0.46$ ($p < 0.0001$), for C2 $r = 0.04$ ($p = 0.3$); for the slow condition for C1 $r = 0.34$ ($p < 0.0001$) and for C2 $r = 0.05$ ($p = 0.27$). The correlations confirm the rate-independent position effect in the durational control of C1, C2.

The presence of a correlation between closure formation and plateau duration for C1, but the lack of such a correlation for C2 may possibly point towards possible trading relations in the durational control of C1 and C2. Conceivably, durational variation within C1 may be compensated for in the closure formation of C2 as response to auditory requirements, for instance in terms of a certain plateau lag between the two clusters. Such a state of affairs would be a first hint for auditory cue recoverability generally shaping the flexibility of articulatory timing. This in turn would suggest that while there generally is some degree of co-variation in total duration between C1 and C2, the covariation might not be equally distributed across all phases of the movement cycle. To test for this possibility, we correlate as a next step the total duration of C1 as the sum of constriction formation and plateau duration, with closure formation duration of C2 and with plateau duration of C2. The expectation is that C2 constriction formation, but not plateau duration, would vary with total duration of C1. Fig. 4 gives the relevant scatterplots. There is a weak positive correlation between constriction formation duration of C2 and C1 total duration (fast: $r = 0.12$, $p = 0.008$; slow: $r = 0.19$, $p < 0.0001$), but this does not hold in the same way for plateau duration of C2 (fast: $r = -0.14$, $p = 0.0014$, $r = -0.03$, $p = 0.52$). There is

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

Table 3
Average constriction formation duration (CLO), plateau duration (PLAT), and total duration in ms for C1 and C2 by rate. Standard deviations are in parentheses.

Rate	C1 CLO	C2 CLO	C1 PLAT	C2 PLAT	C1 Total (CLO + PLAT)	C2 Total (CLO + PLAT)
Slow	72 (23)	77 (21)	65 (32)	58 (25)	138 (45)	135 (33)
Fast	62 (19)	69 (15)	53 (23)	44 (16)	115 (36)	113 (23)
Total mean	67 (22)	73 (19)	57 (28)	51 (22)	127 (43)	124 (31)

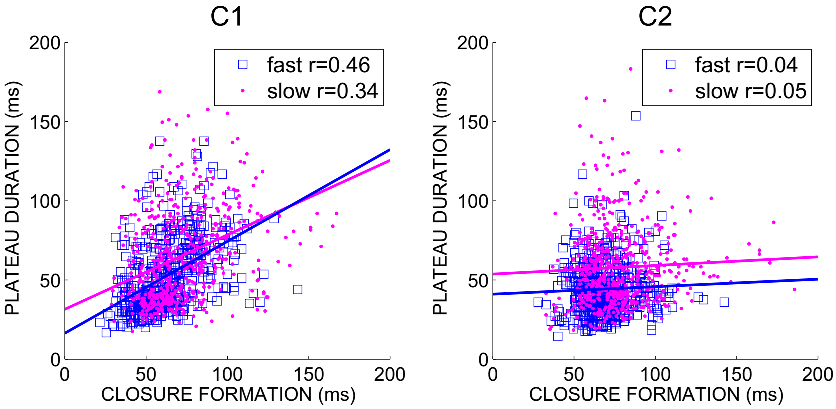

Fig. 3. Scatterplot of plateau duration as a function of closure formation duration for C1 and C2 by rate. The graphs reveal a position effect in that closure formation and plateau duration are modestly correlated for C1, there is no such correlation for C2.

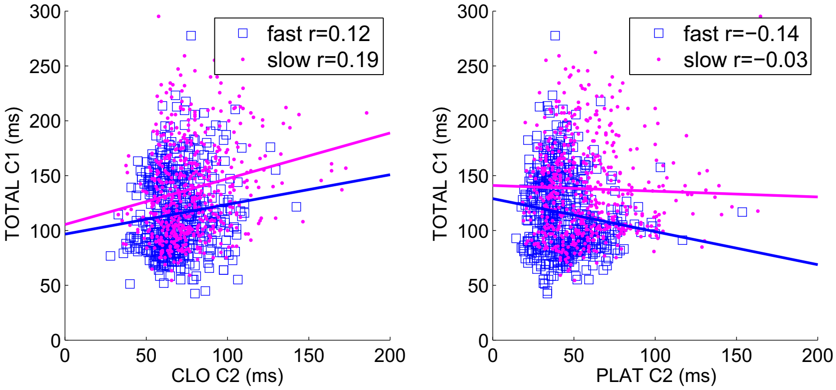

Fig. 4. Scatterplot of closure formation duration (left) and plateau duration (right) of C2 against total duration of C1. The expectation is that C2 constriction formation, but not plateau duration, would vary with total duration of C1. There is only weak support for this expectation.

no correlation at the slow rate and a weak negative correlation at the fast rate. In sum, while there is evidence for a positional effect on consonant duration, there is no strong evidence for the durational variation of C2 constriction formation being correlated with durational variation of C1. The timing between the two consonants will be considered in more detail in the next section, in which we also turn to our main question of whether relative timing varies as a function of the auditory cue profile of a cluster and/or cluster frequency.

3.3. Consonant timing: plateau lag and onset lag as a function of auditory cue profile

In this section we consider how the timing between the two cluster members differs as a function of speaking rate and a possible interaction with the cue profile of a cluster. As laid out in the Introduction, the particulars of Russian cluster phonotactics have led us to hypothesize that C1-C2 plateau lag might globally vary relatively little as a function of speaking rate. This analysis therefore speaks to the general hypothesis laid out in the Introduction that the existence of auditorily ‘difficult’ clusters may hinge on a relatively low degree of C-C overlap and a concomitant immunity to rate changes. The uniform shortening of plateau durations with rate evident in Table 3 already gave some indication that this conjecture may not find support in the data.

Fig. 5 plots the plateau lag results for absolute (Fig. 5a) and normalized (Fig. 5b) durations, as well as onset lag (Fig. 5c) by cue group and rate. Looking at plateau lag (Fig. 5a) first, it is apparent that plateau lag changes with rate for all groups, against our initially formulated conjectures. Note that this is true in absolute as well as in proportional time (Fig. 5b): When plateau lag is computed as a proportion of C1 duration (Fig. 1c), there still is a shortening of the inter-plateau interval at the fast rate. Thus it is neither the case that Russian clusters generally, nor those from the poor auditory cue group resist rate changes. At the same time, it is noteworthy that there is a positive plateau lag across the data (> 0 in absolute and

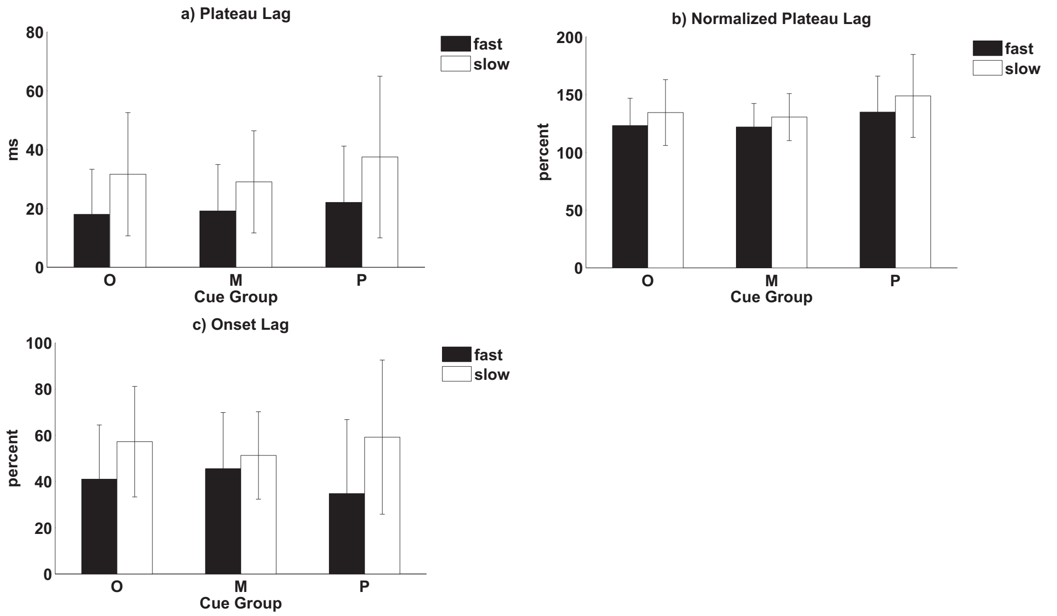

Fig. 5. (a) Absolute and (b) normalized plateau lag durations as well as (c) onset lag by Cue Group (O = optimal, M = medium, P = poor) and Rate. Whiskers indicate ±1 SD computed as average of the by-subject standard deviations. The data support a rate effect across the board for all three auditory cue profile groups.

>100 in normalized time), meaning the constriction plateaus of the two consonants do not overlap at either rate, a point we will come back to in the Discussion. The statistical results are as follows: A linear mixed model with main factors Cue Group and Rate for the raw plateau lag fails to render a significant interaction ($\chi^2(2) = 3$, $p = 0.22$), gives a significant effect of Rate ($\chi^2(1) = 15.73$, $p = 0.00007$) but not of Cue Group ($\chi^2(2) = 2.4$, $p = 0.3$). For the normalized lag, the results pattern the same with a non-significant interaction ($\chi^2(2) = 1.85$, $p = 0.39$), an effect of Rate ($\chi^2(1) = 16.2$, $p = 0.00005$) but not of Cue Group ($\chi^2(2) = 3.59$, $p = 0.17$). In sum, plateau lag shortens with fast rate for all cluster groups in both absolute and proportional time. There is neither an effect of Cue Group nor an interaction of Cue Group with Rate. Estimated effect size for going from fast to slow Rate in the full model is in absolute time 9.56 ms (se ±2.94), and in proportional time 8% (se ±3.17).

We now turn to the onset lag measure; the results are in Fig. 5c. Recall that this measure essentially quantifies the percentage of C1's duration occurring prior to movement onset of C2 (see Methods, particularly Fig. 1c). Overlap values are consistently smaller for the fast rate. This is in agreement with the increase of anticipatory coarticulation at fast speaking rates reported in previous research (see Section 1.1). However, in contrast to plateau lag the cue groups do not seem to be affected by rate to an equal degree with the medium group being least amenable to change. A linear mixed model with fixed factors Rate (slow, fast) and Cue Group (optimal, medium, poor) confirms this in terms of a significant interaction ($\chi^2(2) = 22.39$, $p < 0.00002$), a significant effect of Rate ($\chi^2(1) = 14.43$, $p < 0.00014$), but no significant effect of Cue Group ($\chi^2(2) < 1$, $p = 0.99$). The significant interaction is due to a significant difference between Rates for the optimal and poor, but not the medium Cue Group (Tukey posthoc, $p < 0.0001$ for optimal and poor; $p = 0.64$ for medium). Fig. 6 gives the estimated effects for the interaction. Against all expectations, the rate separation is strongest for the poor cue group.

In sum, both plateau lag and onset lag significantly decrease with increasing rate. This indicates that overall, a faster rate is concomitant with a higher degree of coarticulation, confirming for consonant clusters earlier research on rate effects in speech production (Section 1.1.). Speaking rate interacts with cue group in that the medium cue robustness group does not display variation in onset lag as a function of rate. There is neither an interaction effect nor an effect of Cue Group for plateau lag. For a better understanding of these results, Fig. 7 visualizes normalized C1-C2 timing by cue group. Looking at the optimal (leftmost column) and poor cue (rightmost column) groups first, we can clearly see that the increased coarticulation between C1 and C2 is caused by a proportional increase in constriction formation duration of C2 at the fast rate. For the medium profile clusters only (middle column), we see in essence invariant relative timing across rates.

Our initial hypothesis was that Russian clusters may generally show little flexibility under speech rate changes due to the wide array of phonotactic possibilities for onsets in this language. This could clearly be disconfirmed: there is a strong effect of rate in all of our measures in terms of absolute as well as proportional time. The individual cluster members both shorten with rate, and there is a global increase in coarticulation at the fast rate caused mainly by a proportional increase in C2 constriction formation duration. At a more detailed level, the research question targets a possible interaction between auditory cue robustness and speech rate. Against expectations, there is neither a significant interaction nor a significant effect of cue group. All groups shorten their plateau lag equally with rate. Nonetheless plateau lag remains above zero in absolute time (>100 in normalized time), meaning the constriction target plateaus of the two consonants generally do not overlap. The results for plateau lag contrast with those of onset

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

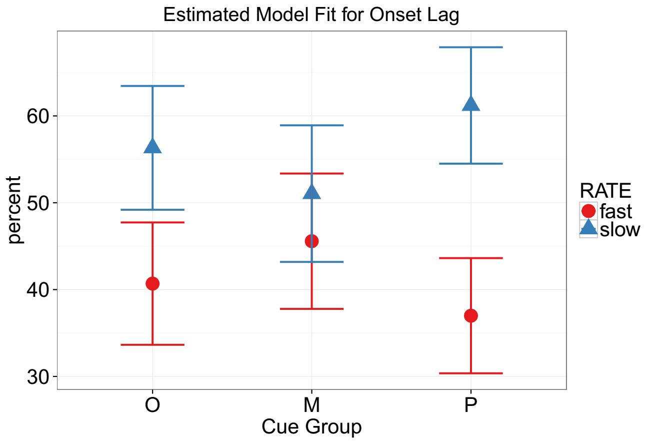

Fig. 6. Estimated model fit for the Cue Group * Rate interaction for dependent variable onset lag. Whiskers indicate 95% confidence intervals. Rate separation is strongest for the poor cue group against the predictions of the auditory cue recoverability hypothesis.

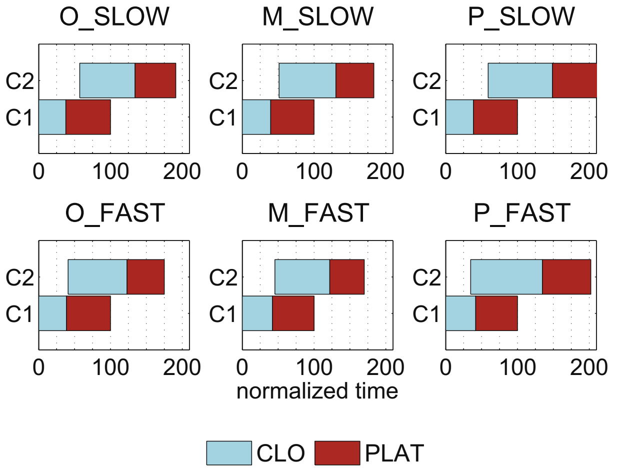

Fig. 7. Bar graphs of average relative C1-C2 timing by profile group (left-to-right: optimal (O), medium (M), poor (P)) and rate (top: slow, bottom: fast). All time points are normalized to the duration of C1 (see Methods). The light (cyan) bars represent constriction formation duration intervals, the dark (red) bars plateau duration. The data are cross-subject averages. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lag in terms of a significant interaction between speech rate flexibility and cue group for the latter; thus, for the onset lag measure, only the medium cue group (/lb, lg, ml/) is more or less invariant in its relative timing. The most pronounced rate effect is evident for the poor cue group. This result remains unsatisfactory, since we would, based on auditory cue robustness considerations alone, expect an ordering of optimal $\geq$ medial $\geq$ poor with respect to effect size. An exemplary visualization of the clusters /bl, lb, kt, tk/ in Fig. 8 underscores that auditory cue robustness alone is not the best predictor of the results (corresponding plots are given for all clusters in Fig. 13 in the Appendix). Comparing slow and fast for cluster /bl/, an optimal cue cluster, in the top and bottom left-most graphs of Fig. 8, we clearly see the increase in onset lag: C2 begins its movement earlier relative to C1 in the fast compared to the slow condition. For /lb/, a representative of the medium group, there is virtually no change. The poor cue group, here represented by /kt, tk/, patterned statistically with the optimal profile group (here represented by /bl/). /kt/ indeed behaves like /bl/, but observe how /tk/ is rather similar to /lb/ than to /kt/. This means that even between two stop + stop clusters there is very different behavior in terms of onset lag. This in turn suggests that while we observe a statistical result by cue group, auditory cue robustness is not the main predictor of a cluster's behavior under speech rate manipulation, neither for plateau lag nor for onset lag. We now consider frequency as an alternative grouping factor. In many cases, high frequency and phonotactic preference (high degree of cue robustness) are inseparable and it is very difficult to examine these two factors independently of each other. Russian allows us to circumvent this confound to a certain degree, since the grouping of clusters by cue profile and by frequency is not isomorphic, even though of course we are not able to de-correlate these two effects completely in a crossed design. Nonetheless, we can gain some first insights here by examining whether there is an interaction of frequency with rate.

3.4. Frequency effects

As argued in Section 2.3, we operationally created three frequency bins for data analysis in order to ensure comparability between the auditory cue analysis section and present analyses. For our key findings, we also report results for frequency as a continuous predictor. Fig. 9 gives the means for absolute (a) and normalized (b) plateau lag and onset lag (c) for a grouping by frequency bin and rate.

As for the auditory cue group results, we see a consistent rate effect across the three groups for plateau lag, but here the effect seems attenuated for the low frequency group. A differential rate effect between the high and low frequency groups is quite pronounced for the onset lag measure. Linear mixed models (fixed effects: Rate (slow, fast), Frequency Group (high, mid, low)) confirm these impressions statistically: For raw plateau lag values, the interaction just about misses significance ($\chi^2(2) = 5.95$, $p = 0.051$). Rate is significant at $\chi^2(1) = 15.74$, $p = 0.00007$, but Frequency Group is not ($\chi^2(2) = 0.32$, $p = 0.85$). For the normalized plateau lags, the interaction does reach significance ($\chi^2(2) = 10.31$, $p = 0.006$). Rate is again significant at $\chi^2(1) = 16.22$, $p = 0.00006$, but not Frequency Group ($\chi^2(2) = 0.70$, $p = 0.71$). A Tukey posthoc underscores that the significant interaction is due to the low frequency group not changing plateau lag significantly with rate (low $p = 0.33$; mid $p = 0.006$; high $p > 0.0001$). In terms of onset lag, there is again a significant interaction ($\chi^2(2) = 19.15$, $p = 0.00007$). There is also a significant effect of Rate ($\chi^2(1) = 14.38$, $p < 0.00015$) and, in contrast to plateau lag, an effect of Frequency Group ($\chi^2(2) = 13.31$, $p = 0.0013$). Posthoc tests confirm that the interaction is due to a significant rate difference for the high and mid ($p < 0.0001$) groups only (low: $p = 0.44$). The Frequency Group effect is due to the low group having significantly longer onset lags than both the high ($p < 0.0001$) and mid ($p = 0.007$) groups, with the latter two not differing from each other ($p = 0.37$). We also ran a mixed linear model on onset lag with the mean-centered cluster frequency scores as linear predictor; the Rate variable was centered on 0 with a distance of 1. Fig. 10 gives the model fit. Results are in complete agreement with the the binned frequency analysis; onset lag changes with speech rate to a

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

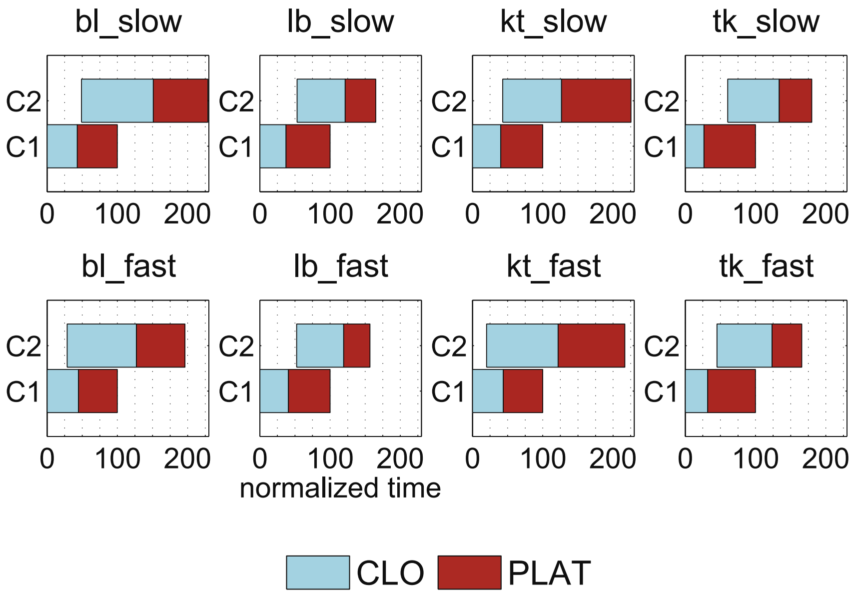

Fig. 8. Bar graphs of average relative C1-C2 timing by cluster. The light (blue) bars show closure formation the dark (red) bars plateau duration for each consonant in normalized time (see Methods). Slow rate is at the top, fast rate at the bottom. The graphs represent across subject averages. /bl/ represents the optimal cue group, /lb/ the medium, and /kt, tk/ the poor cue group. The similarity in timing between /kt/ and /bl/ on the one hand and /tk/ and /lb/ on the other underscores that cue robustness is not the best predictor of the results (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

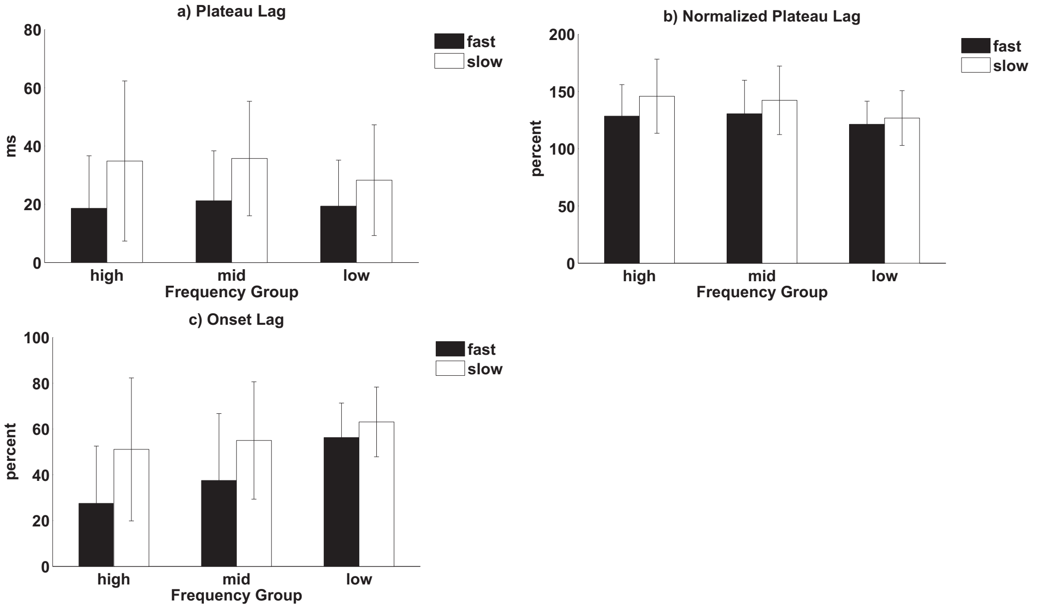

Fig. 9. (a) Absolute plateau lag, (b) normalized plateau lag, (c) onset lag by Frequency Group and Rate. Whiskers indicate ±1 SD computed as average of the by-subject standard deviations. Normalized plateau lag and the onset lag measure show a differential rate effect between high and low frequency groups.

significantly higher degree for the high compared to the low frequency clusters. A step increase in frequency renders an estimated −3.8% (se ±0.8, $t = -4.5$) decrease in onset lag. For the fast rate, the decrease per frequency step is −5%, for the slow rate it amounts to −2.4%.

In sum, onset lag decreases with increasing rate, albeit not for the low frequency group. We conclude that the results confirm our predictions in that there is an interaction between frequency group and rate, with the effect being more pronounced with increasing cluster frequency.

There is a global rate effect across the board for all measures. However, rate effects do not impact all conditions equally, and the interaction of rate with either cue robustness or frequency seems to be stronger in a particular part of the cluster. Any interactions present were most consistently apparent in onset lag, i.e. the relative movement onset of C2 within C1. Since there is to our knowledge next to no published work on how speech rate effects are manifest in consonant clusters we take the time here to consider in more detail the non-uniform rate scaling that has become apparent in our analyses

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

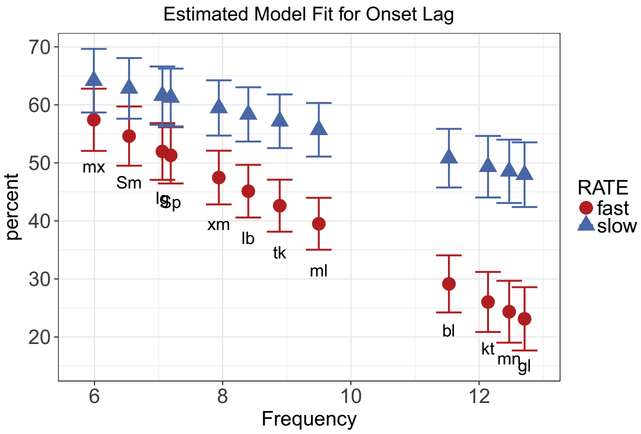

Fig. 10. Estimated model fit for the Frequency*Rate interaction for dependent variable onset lag. Whiskers indicate 95% confidence intervals. The separation between fast and slow rate is stronger for the higher frequency clusters.

Table 4
Statistical results for changes to constriction formation duration (CLO) or plateau duration (PLAT) for C1, C2 in relative cluster time.

		Rate * Frequency $\chi^2(1)$, p
C1	CLO	0.42, $p = .43$
	PLAT	2.5, $p = .11$
C2	CLO	13.4, $p = .00026$
	PLAT	.16, $p = .69$

thus far, and return in the Discussion to the ramifications of our findings for models of the durational control of gestures.

In order to gain a comprehensive understanding of how rate effects are manifest throughout the cluster, we ran a series of statistical models on plateau duration and constriction formation duration for both C1, C2 in normalized time. Here we normalize for cluster duration (defined as time point of C1 peak velocity of constriction formation to release of C2, see Methods) in order to be able to assess the relative contribution of both C1 and C2 to proportional duration changes within the cluster, and where exactly changes in relative timing between C1 and C2 are localized. In order to limit the number of statistical tests, we restrict ourselves here to the interaction of Rate with Frequency, since this interaction is the main interest here and the most consistently interpretable pattern in our results. The statistical results are summarized in Table 4. The model specified an interaction between Rate (slow, fast) and Frequency (each cluster's mean-centered frequency value). Since this analysis is not in direct comparison with the auditory cue group analyses, we only used the actual, mean-centered cluster frequency scores as linear predictor and did not also group clusters into frequency bins. For interpretability, the Rate variable was also centered on 0 with a distance of 1.

For graphical display purposes, we computed for each normalized time measurement the average of the slow condition on a by-speaker by-cluster basis and subtracted from this mean each token of the fast condition. This gives us, for each token of the fast condition, a distance value quantifying the change under rate. The subtraction was always mean (slow) – fast, thus positive values indicate that the slow condition had a larger measurement value compared to the fast condition. Fig. 11 displays the distance values.

There is no significant interaction in Table 4 for normalized plateau duration, neither for C1 nor for C2 (top row graphs of Fig. 11). For constriction formation duration, we see a relative increase in proportional duration in the fast condition (medians below zero) for both C1 and C2. For C1, this effect is relatively uniform across frequency groups, but for C2 the high frequency group shows the largest difference values. This is in accordance with the results in Table 4, the only significant interaction arising for constriction formation of C2. This means that the increase in overlap in the high frequency condition revealed in our previous analyses is due to an earlier movement onset of C2 within C1, concomitant with proportionally longer constriction formation duration of C2. Interestingly, this does not hold for plateau duration of C2, underscoring that rate-scaling is not only non-uniform throughout the cluster, but even non-uniform within C2. Fig. 12 gives the estimated linear model fit for the significant interaction between Rate and Frequency for C2 closure formation duration. Effect estimates of the model render a distance between the slow and fast rate of 6% (se ±1.23, $t = -4.75$). A step increase in frequency renders an estimated 0.7% (se ±0.5, $t = 1.4$) lengthening of proportional C2 constriction formation duration. For the fast rate, the lengthening effect per frequency step is 1.15%, for the slow rate it amounts to only 0.25% (interaction $t = -3.67$). This confirms that the significant interaction is due to a decreasing rate effect with decreasing cluster frequency.

In sum, the interaction effect between rate and frequency is carried by C2 and within C2 mostly by the part during which C1 and C2 co-exist, i.e. C2 constriction formation. High frequency clusters show a larger amount of coarticulatory integration which is achieved by means of a proportionally longer constriction formation duration of C2.

4. Discussion

The goal of our study was to investigate the durational characteristics of Russian consonant clusters, how they vary with speaking rate, and whether rate interacts with auditory cue robustness and/or frequency differences between clusters. Thereby we took up a proposal by Wright (1996) and subsequently Henke et al. (2012) on the conditions under which typologically unusual clusters can be stabilized within a language's phonotactic inventory. The central tenet of that proposal is that clusters which under a canonical coarticulation pattern would suffer from poor auditory cue robustness may exist in languages conditional on a low degree of consonant coarticulation. This 'minimal coarticulation constraint' for these clusters will subsist in conditions under which coarticulation usually increases, such as increasing speech rate. Our present focus on Russian was therefore motivated by the range of clusters allowed in this language together with reports from previous research that Russian consonant sequences show a low degree of plateau overlap and, compared to English, only limitedly vary in duration as a function of prosodic affiliation. Wright (1996) had presented initial evidence that the existence of those clusters within a language is conditional on a low overlap pattern (cf. also Chitoran, 2016; Pouplier & Beňuš, 2011) and that this entails a lesser susceptibility of these clusters to changing their timing characteristics with rate. We also

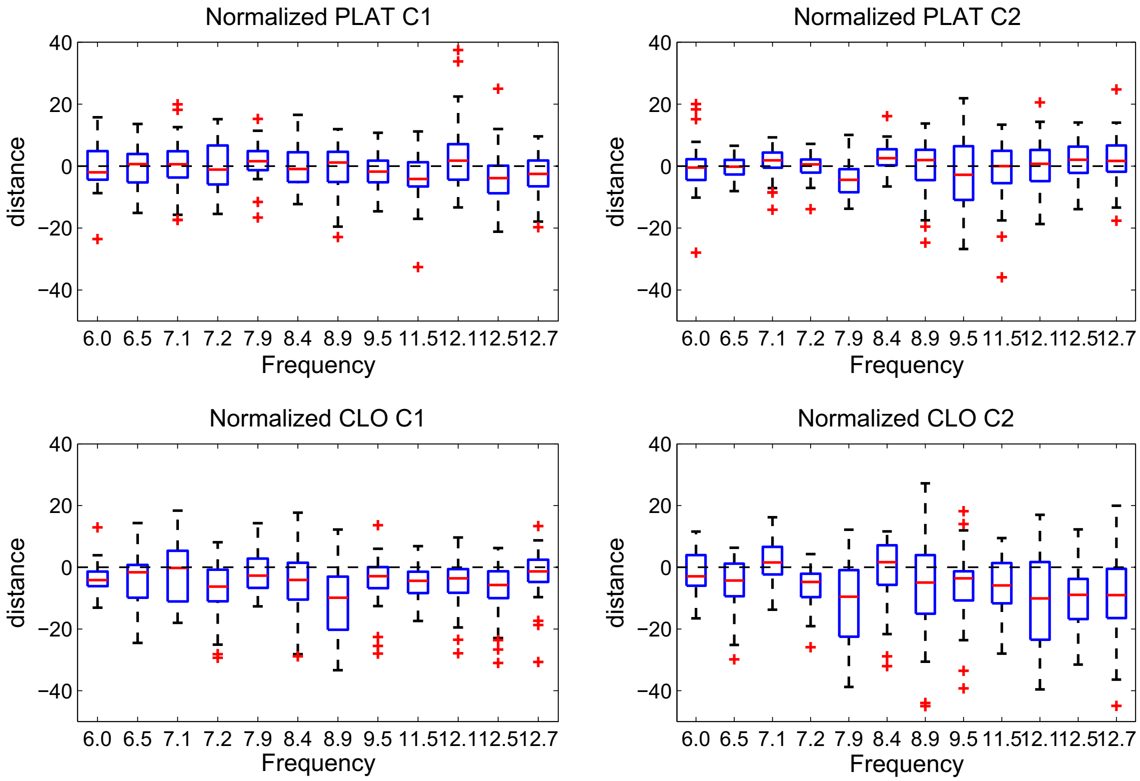

Fig. 11. Distance (percent) to the slow condition for normalized plateau duration (PLAT, top row) and normalized constriction formation duration (CLO, bottom row) of C1 (left column) and C2 (right column) by frequency. The data support a Rate*Frequency interaction for closure formation duration of C2 only (bottom right graph).

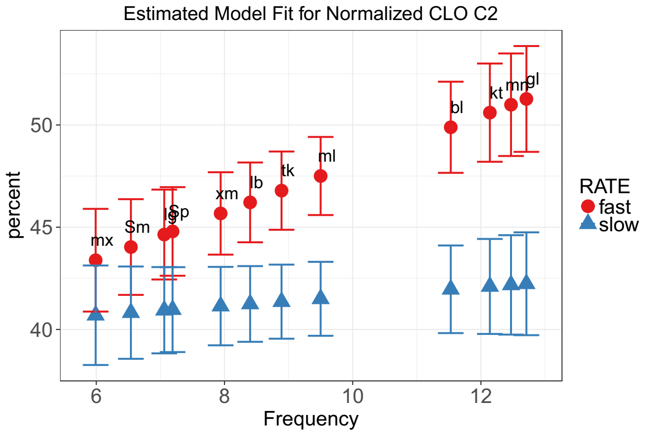

Fig. 12. Linear model estimate of the Frequency by Rate interaction effect for normalized constriction formation duration (CLO) of C2 (proportional to cluster time). Whiskers indicate 95% confidence intervals.

asked whether cluster frequency may entail differential rate flexibility.

Our results firstly confirm that there are pronounced rate effects in Russian onset clusters. Initially we had hypothesized that Russian might be resistant to implementing rate effects in clusters globally as part of a language-specific 'low overlap' coarticulatory setting. This was not confirmed: in absolute time, all cluster components shorten with rate, and also globally the relative timing between the consonants changes. Thereby durational effects are not distributed uniformly throughout the consonant cluster, a point we shall return to below. When grouping clusters according to their presumed auditory cue robustness, we found a statistically significant interaction in that the medium robustness group had no increasing onset lag with increasing rate. However, the poor auditory cue robustness group which was expected to be most resistant to rate changes instead showed the most pronounced rate effect on onset lag and statistically aligned with the optimal cue group, making a straightforward interpretation of the results based on cue robustness difficult. When considering cluster frequency, there likewise was an interaction with rate, but here a much more coherent picture emerged: Low frequency clusters did not increase anticipatory coarticulation with rate and they showed in proportional time little flexibility in constriction formation duration of C2. This contrasts with high frequency clusters which show an earlier movement onset of C2 within C1 with a concomitant increase in constriction formation duration of C2.

Overall, the auditory cue robustness hypothesis is not supported by our results. Wright (1996) used Tsou stop + stop clusters to argue that clusters composed of sounds with poor internal cues are resistant to rate changes since their very existence in the language is conditional on a lesser degree of overlap within the language compared to optimal cue clusters within the same language. This is not what we found; in our data there was neither an effect of cue group nor a meaningful interaction of cue group and rate. Nonetheless, we note that even though all cue groups shortened their plateau lag with rate, plateau lag remained largely above zero (>100 in normalized time) across all clusters; accordingly one might argue that plateau lags in Russian as a whole are never shortened to a degree that might endanger auditory cues. Since Russian has a relatively large plateau lag to begin with, rate

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

conditioned shortening and an increased anticipatory coarticulation might still leave enough room for cue recoverability. This in turn could be connected to our initial hypothesis that the characteristics of speech rate changes may be part of a language-specific coarticulatory setting, but to evaluate this in detail a cross-linguistic study will be necessary. Whether more extreme rate variation would reveal differential plateau lag stability as a function of cue robustness in Russian is unknown and would have to be tested in future research. In this context, we would also like to caution against directly setting articulatory lag measures in relation to the hypothesized perceptibility of clusters: it is important to keep in mind that the articulatory release, especially for stops, is not synchronous with the acoustic release. Due to the viscosity of the articulators in conjunction with stops being characterized by a virtual target beyond the point of contact (e.g., Löfqvist and Gracco (1997)), the articulatory release may considerably precede the acoustic release. Moreover, articulatory landmarks obviously are only very limitedly informative about intraoral pressure dynamics which are however an important factor for the resulting acoustics and ultimately percept. Generally, very little is known about how exactly the type of particular timing measures reported here and in a multitude of other studies relate to the degree of perceptual recoverability, even though a rather direct relationship is generally (implicitly) assumed. A detailed modelling study of the relationship between the articulatory dynamics of a cluster and the resulting acoustic patterns is called for in this context.

A general point to consider here is how auditory cue robustness would come to influence articulatory timing patterns at all. Even though this relationship is invoked in many publications (including our own), there are few proposals on the pathways via which such an interaction might actually arise. Tilsen (2016) proposes that the skill required for producing multiple articulatory events in an overlapping fashion (as is required for onset clusters) is characterized by a reliance on state-feedback control rather than external feedback, and learning complex coordination patterns is tantamount to shifting from a reliance on external feedback to a reliance on internal feedback (feedback internalization). It is only under state-feedback control that gestures can be co-selected during planning and hence be produced in an overlapping fashion. He specifically mentions that perceptual recoverability forces (among a wealth of other factors) block a shift to state feedback control and thus prevent the shift from sequential to highly coarticulated productions characteristic for highly overlearned patterns otherwise. But note that none of our measures yielded a significant effect of cue group, nor was there a meaningful interaction with rate, i.e. our study generally found no robust support for a covariation between perceptual recoverability and degree of overlap. It remains to be seen whether other studies can provide evidence for perceptual recoverability conditioning a behavioral patterning parallel to low frequency (i.e. a lower degree of overlearnedness).

In further defense of the auditory cue hypothesis, one might say that our poor auditory cue group was not limited to stop + stop clusters, but recall the very different patterning of /tk/ and /kt/ in Fig. 8. (In fact, frequency provides a much better alignment between /bl, kt/ and /lb, tk/; we give normalized time plots for all clusters in Fig. 13 in the Appendix). An alternative perspective to take on auditory cue robustness might be to argue that clusters with poor auditory cues increase their coarticulation in order to increase the temporal spread of their cues to a greater extent. A similar argument has been considered by Scarborough (2013) in the context of lexical neighborhood effects in speech production. She found that words with a high neighborhood density are coarticulated more than those from a low-density neighborhood. While this reasoning may help understand why the poor cue group would show the greatest increase in coarticulation with rate, auditory cue robustness remains a poor predictor for the overall pattern in the data.

The just mentioned discrepant patterning of /tk, kt/ can be used to highlight another important factor to consider when looking at a range of different clusters as we did in this study. On purely articulatory or biomechanical grounds /tk/ is not a mirror image of /kt/ (neither /lg/ of /gl/, etc.). The movement dynamics of the closure-release formation are very different in these two cases; both articulator and articulator sequence may be confounding factors in our analyses. Also acoustically, overlapping /t, k/ productions may have very different consequences depending on the fine details of the constriction dynamics. Marin, Pouplier, and Harrington (2010) investigated the acoustic consequences of overlapping /t, k/ specifically in the context of speech errors. They found that an overlapping velar constriction has a distorting effect on the burst spectrum of /t/, but vice versa this was not the case. The burst spectrum of a velar is unaffected by an overlapping coronal constriction. While in their (American English) speech error data the constrictions probably overlap to a greater degree than in regular speech, their main finding again drives home the point that clusters differing ‘only’ in the serial order of their segments may differ physiologically and acoustically to a much greater degree than is suggested by the symbolic order reversal. Of course it is impossible to exert experimental control over these cluster intrinsic effects since also manner and voicing differences affect consonant timing (e.g., Bombien & Hoole, 2013; Marin, 2011), and therefore clusters varying according to some independent variable by nature also always vary in their intrinsic properties. Other definitions and metrics of auditory cue robustness are conceivable, and our heuristic grouping into three cue groups has to ultimately remain unsatisfactory given that cue robustness is an inherently graded phenomenon. However, what we can say is that if the reasoning of Henke et al. (2012) on the context-specific nature of cue robustness is used as a tool to categorize our stimuli, there is no support for the hypothesis that cue robustness conditions the speech rate flexibility of a cluster.

When frequency was considered as an alternative grouping factor, a very consistent pattern emerged with the strength of the rate effect decreasing for lower frequencies (Fig. 10, Fig. 12). There is only a small body of work on the effects of frequency on articulation dynamics, but our results are consistent with previous findings that higher frequency leads to an increase in anticipatory coarticulation: There was an effect of frequency for onset lag (but not for plateau lag). One contribution that our study could make is to show that frequency also entails a differential coarticulatory flexibility to rate scaling. High, but not low frequency clusters show a significant increase in coarticulatory integration with rate. Munson (2001) failed to find an interaction between speech rate and

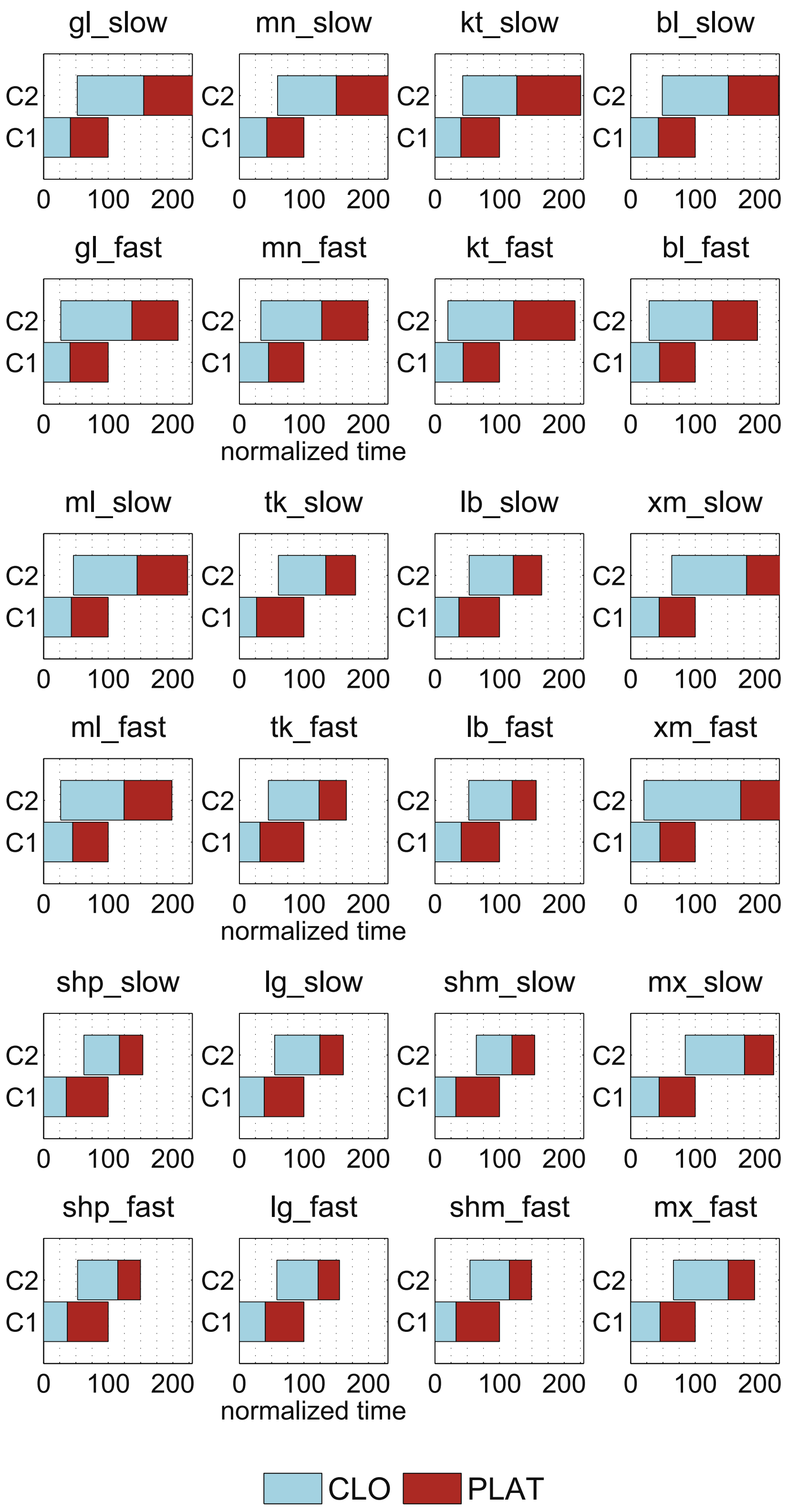

Fig. 13. Normalized time plots for all clusters in order of decreasing frequency. For each cluster, slow rate is at the top, fast rate at the bottom. The light (cyan) bars represent constriction formation duration intervals, the dark (red) bars plateau duration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

frequency in terms of variability. While low-frequency phoneme sequences were in his study articulated with longer and more variable durations than high-frequency sequences, this was the same across speech rates. In our study, it was the proportional time measures that underscored how low frequency entails a reduced articulatory flexibility to contextual changes

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

(here, rate). Consistent with our results, Zellou and Tamminga (2014) found that anticipatory nasal coarticulation is more extensive in high frequency words independently of the types of durational shortening typically associated with reduction.

Again we should consider whether the frequency effects may be confounded by articulator specific patterns. One comparatively well understood factor that conditions differences in coarticulatory flexibility is coarticulation resistance (even though we are not aware of a study looking at the interaction of coarticulation resistance and speech rate). We can thus briefly consider whether the frequency-based ordering of clusters aligns with their global coarticulation resistance. Here we will take the coarticulation resistance of C1 into account since rate scaling induces increased presence of C2 during C1. Understanding the coarticulation resistance of consonants can be quite complex (see Farnetani and Recasens (2010) for an overview), but for present purposes it is reasonable to assume that globally labials are least resistant to coarticulation and sibilants the most. Therefore /bl, mn, ml, mx/ should show the highest degree of rate change and /ʃm, ʃp/ the lowest. This is partially confirmed in that /bl, mn/ are high frequency clusters and /ʃm, ʃp/ are low frequency clusters. /mx/ is the lowest frequency cluster, /ml/ is the 5th highest frequency cluster. Fig. 10 and Fig. 12 show that /mn, bl/ and /ml/ (second, fourth, fifth data points from the right) are part of the linearly decreasing rate effect, /mx/ clearly patterns against the most simple coarticulation resistance predictions. Overall, articulatory effects may play a role, but they do not seem to be the main factor conditioning our results.

Coarticulation is a learnt skill (e.g., Munson, 2001; Noiray, Ménard, & Iskarous, 2013; Zharkova, Hewlett, & Hardcastle, 2011) and our results are in line with a large literature on frequency effects in spoken language suggesting that the role of experience permeates into the mature adult system. This has ramifications for claims that statistical asymmetries in the distribution of phonotactic patterns may be due to inherent, possibly universal difficulty or complexity of these patterns be it on a physiological, perceptual or cognitive level: Our data offer support for the idea that for mature adult systems, complexity cannot be defined in a language-independent manner (Pouplier, 2012, 2015) – even though physiological constraints on production and perception may of course interact with frequency in language acquisition (see Edwards and Beckman (2008), Edwards, Beckman, and Munson (2015) for a discussion of the interaction of phonetic, substantive universals with language-specific frequency effects during phonological development).

In sum, there is no support in our data for the idea that individual clusters within a language may be immune to rate changes due to their auditory profile requiring a lower degree of coarticulation compared to other clusters within that same language. We did find, however, that plateau lag as a whole was less susceptible to change than movement onset of C2. Any differential cluster behavior in our data is best explained on the basis of language-internal frequency rather than cluster typology. Our results overall therefore do not allow us to evaluate any further the question of whether the existence of particular phonotactic patterns in a language implicates a particular global coarticulatory setting: Russian clusters are, despite their typologically unusual phonotactics, subject to durational shortening and increasing coarticulation under increased rate. Carefully controlled cross-linguistic work will have to test the characteristics of rate variation for languages with different coarticulation profiles. If different patterns of rate variation were found, this might still allow for a more general version of the auditory cue robustness hypothesis.

As a final point, we would like to raise the question of how our results could be incorporated into models of duration control. This very basic question is central in particular for discussions on how to conceptualize time and timing in spoken language (Fowler, 1980; Sorensen & Gafos, in press; Turk & Shattuck-Hufnagel, 2014). In our data, rate effects, if occurring, are predominantly carried by the constriction formation of C2 which is not easily accounted in the most common model of durational control which is based on the linear second-order mass-spring model (e.g., Ostry & Munhall, 1985). Thereby duration is not controlled directly but duration and movement endpoint are specified in terms of one or more of the spring constants (stiffness, damping, resting length of the spring). Correspondingly, durational change is brought about indirectly through the specification of the biomechanical parameters of the articulators for a given gesture. Stiffness and damping remain constant for a given gesture, but can be altered between gestures. But there was no correlation between constriction formation and plateau duration for C2 at either rate. Moreover rate effects were mostly local to the constriction formation duration of C2 (cf. Fig. 11). Relatedly it has been observed previously that rate effects do not simply rescale the velocity profile but condition a qualitative change giving evidence for corresponding qualitative differences in underlying movement control (Adams et al., 1993). In particular the linear second order model which has been employed among others by the task-dynamic model (Saltzman & Munhall, 1989) has been criticized as too simplifying given the observation of plateaus (geminates, long vowels), which requires time-variation in damping and stiffness within a gesture (Fuchs, Perrier, & Hartinger, 2011). To more appropriately model the articulatory movement cycle, higher order models have been proposed (Birkholz, Kröger, & Neuschaefer-Rube, 2010), as well as models advocating a separate closure and release gesture for consonants (Browman, 1994; Harrington, Fletcher, & Roberts, 1995; Nam, 2007). The latter would not necessarily predict a correlation between constriction formation and plateau duration. Nonetheless also for these models it remains for now unresolved why we would see a higher degree of durational co-variation for different parts of the movement cycle within C1 but not C2, and why speech rate changes would mostly be local to constriction formation duration of C2. Finally, our results relate to an old discussion about invariant relative timing. While a series of papers by Tuller, Kelso, and colleagues (Kelso, Saltzman, & Tuller, 1986; Kelso & Tuller, 1987; Tuller, Kelso, & Harris, 1982) proposed in the 1980 s under the impetus of a dynamical systems approach to speech that there was invariant relative timing in speech production, subsequent work cast doubt on the invariance hypothesis (Nittrouer et al., 1988; Shaiman et al., 1995). Which factors may condition changes to relative timing is currently not understood. The dynamical systems approach has been tightly linked to the idea of spoken language as an acquired skill (Turvey, 1990) which enables one to flexibly adapt one's

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

actions to a plethora of different contexts. Our present results suggest that conditional flexibility of relative timing patterns themselves, as witnessed by speech rate changes, is part and parcel of this skill, reminiscent of Lindblom's idea of the gaits of speech (Lindblom, 1989; Pouplier, 2012).

Acknowledgements

We thank our RAs Susanne Waltl, Anna Ratzinger, and Alexander Belyy for assistance with speaker recruitment, data recording, and segmentation. Tom Lentz provided helpful advice on statistical issues. Work supported by the ERC under the EU's 7th Framework Programme (FP/2007-2013)/Grant Agreement n. 283349-SCSPL.

Appendix

See Fig. 13.

References

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research, 36*, 41–54.

Agwuele, A., Sussman, H., & Lindblom, B. (2008). The effect of speaking rate on consonant vowel coarticulation. *Phonetica, 65*, 194–209.

Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language, 88*, 148–159.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America, 119*, 3048–3058.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 667*, 1–48.

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.

Berry, J. (2011). Speaking rate effects on normal aspects of articulation: Outcomes and issues. *Perspectives on Speech Science and Orofacial Disorders, 21*, 15–26.

Birkholz, P., Kröger, B., & Neuschaefer-Rube, C. (2010). Model-based reproduction of articulatory trajectories for consonant-vowel sequences. *IEEE Transactions on Audio, Speech, and Language Processing, 19*, 1422–1433.

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*, 539–550.

Browman, C. (1994). Lip aperture and consonant releases. In P. Keating (Ed.). *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III* (pp. 331–353). Cambridge: Cambridge University Press.

Byrd, D., & Tan, C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24*, 263–282.

Chitoran, I. (2016). Relating the sonortiy hierarchy to articulatory timing patterns. A cross-linguistic perspective. In M. J. Ball & N. Müller (Eds.). *Challenging Sonority: Cross-linguistic Evidence* (pp. 45–62). Equinox.

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.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I* (pp. 283–333). Cambridge: Cambridge University Press.

Davidson, L., & Roon, K. (2008). Durational correlates for differentiating consonant sequences in Russian. *Journal of the International Phonetic Association, 28*, 137–165.

Edwards, J., & Beckman, M. (2008). Some cross-linguistic evidene for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development, 4*, 122–156.

Edwards, J., Beckman, M., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*, 421–436.

Edwards, J., Beckman, M., & Munson, B. (2015). Frequency effects in phonological acquisition. *Journal of Child Language, 42*, 306–311.

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

Flemming, E. (2011). La grammaire de la coarticulation. In M. Embarki & C. Dodane (Eds.). *La Coarticulation. Des indices à la répresentation* (pp. 189–211). Paris: L'Harmattan.

Flemming, E., Ladefoged, P., & Thomason, S. (2008). The phonetic structures of Montana Salish. *Journal of Phonetics, 36*, 465–491.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113–133.

Fox, J. (2003). Effects displays in R for generalised linear models. *Journal of Statistical Software, 8*, 1–27.

Fuchs, S., Perrier, P., & Hartinger, M. (2011). A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model. *Journal of Speech, Language, and Hearing Research, 54*.

Gahl, S. (2008). "Thyme" and "time" are not homophonous. Word durations in spontaneous speech. *Language, 84*, 474–496.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America, 63*, 223–230.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica, 38*, 148–158.

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition, 107*, 1155–1164.

Harrington, J., Fletcher, J., & Roberts, C. (1995). Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics, 23*, 305–322.

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.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous interference in general parametric models. *Biometrical Journal, 50*, 346–363.

Jaeger, M., & Hoole, P. (2011). Articulatory factors influencing regressive place assimilation across word-boundaries in German. *Journal of Phonetics, 39*, 413–428.

Jurafsky, D., Bell, A., Gregory, M. L., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229–254). Amsterdam: Benjamins.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14*, 29–59.

Kelso, J. A. S., & Tuller, B. (1987). Intrinsic time in speech production: Theory, methodology and preliminary observations. In E. Keller & M. Gopnik (Eds.), *Motor Sensory Processes of Language* (pp. 203–222). Hillsdale, NJ: Erlbaum.

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

Lin, S., Beddor, P. S., & Coetzee, A. W. (2014). Gestural reduction, lexical frequency, and sound change: A study of postvocalic /l/. *Laboratory Phonology, 5*, 9–36.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1773–1781.

Lindblom, B. (1989). Phonetic invariance and the adaptive nature of speech. In B. Elsendoorn & H. Bouma (Eds.), *Working Models of Human Perception* (pp. 139–173). London: Academic Press.

Löfqvist, A., & Gracco, V. (1997). Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech Language and Hearing Research, 40*, 877–893.

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.

Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America, 88*, 1286–1298.

Marin, S. (2011). Organization of complex onsets in Romanian. In Y. Laprie & I. Steiner (Eds.). *Proceedings of the 9th International Seminar on Speech Production. Montréal*.

Marin, S., & Pouplier, M. (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics, 42*, 24–36.

Marin, S., Pouplier, M., & Harrington, J. (2010). Acoustic consequences of articulatory variability during productions of /t/ and /k/. *Journal of the Acoustical Society of America, 127*, 445–461.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The Cognitive Representation of Speech* (pp. 415–420). New York: North Holland.

Mok, P. (2010). Language-specific realizations of syllable structure and vowel-to-vowel coarticulation. *Journal of the Acoustical Society of America, 128*, 1346–1356.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research, 44*, 778–792.

Munson, B., & Solomon, N. P. (2004). The influence of neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*, 1048–1058.

Nam, H. (2007). Articulatory modeling of consonant release gesture. *Proceedings of the 16th International Congress of the Phonetic Sciences, Saarbrücken, Germany* (pp. 625–628).

Nelson, W. L., Perkell, J. S., & Westbury, J. R. (1984). Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations. *Journal of the Acoustical Society of America, 75*, 945–951.

Nittrouer, S. (1991). Phase relations of jaw and tongue tip movements in the production of VCV utterances. *Journal of the Acoustical Society of America, 90*, 1806–1815.

Nittrouer, S., Munhall, K. G., Kelso, J. A. S., & Tuller, B. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America, 84*, 1653–1661.

Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America, 133*, 444–452.

Ohala, J., & Kawasaki-Fukumori, H. (1997). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In S. Eliasson & E. H. Jahr (Eds.), *Language and Its Ecology: Essays in Memory of Einar Haugen* (pp. 343–365). Berlin: Mouton de Gruyter.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151–168.

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006

Ostry, D., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America, 77*, 640–648.

Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Journal of Laboratory Phonology, 2*, 243–273.

Pouplier, M., Marin, S., & Kochetov, A. (2015). Durational characteristics and timing patterns of Russian onset clusters at two speaking rates. *Proceedings of Interspeech 2015, Dresden, Germany* (pp. 2679–2683).

Pouplier, M. (2012). The gaits of speech: re-examining the role of articulatory effort in spoken language. In M.-J. Solé & D. Recasens (Eds.). *Sound Change* (pp. 147–164). John Benjamins.

Pouplier, M. (2015). Between 'whims of fashion' and 'phonetic law': Performance constraints in speech production in the face of linguistic diversity. In M. Vayra, C. Avesani, & F. Tamburini (Eds.). *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destruttutrazione della strutura sonora del linguagguo/Language acquisition and language loss. Acquisition, change and disorders of the language sound structure* (pp. 41–57). Milano: AISV.

R Core Team (2015). R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Redford, M. (2007). Word-internal versus word-peripheral consonantal duration patterns in three languages. *Journal of the Acoustical Society of America, 121*, 1665–1678.

Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333–382.

Scarborough, R. A. (2012). Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua, 122*, 164–176.

Scarborough, R. A. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics, 41*, 491–508.

Shaiman, S., Saltzman, E. L., & Tuller, B. (1995). Timing relationships of the upper lip and jaw across changes in speaking rate. *Journal of Phonetics, 23*, 119–128.

Sharoff, S. (2005). Russian Internet Corpus. accessed on 30.06.2014 http://corpus.leeds.ac.uk/list.html.

Sharoff, S. (2006). Creating general-purpose corpora using automated search engine queries. In M. Baroni & S. Bernardini (Eds.). *Wacky! Working papers on the Web as corpus*. Bologna GEDIT.

Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., & McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research, 104*, 493–501.

Sorensen, T., & Gafos, A. (in press). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology, 28*(4), 188–215.

Stephenson, L. (2003). An EPG study of repetition and lexical frequency effects in alveolar to velar assimilation. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (pp. 1891–1894).

Tasko, S. M., & Westbury, J. R. (2004). Speed-curvature relations for speech-related articulatory movement. *Journal of Phonetics, 32*, 65–80.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics, 55*, 53–77.

Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. *Interspeech*, 1302–1306.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 460–472.

Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B, 369*, 20130395.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45*, 938–953.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 514–529.

Wright, R. A. (1996). Consonant clusters and cue preservation in Tsou. PhD Dissertation UCLA.

Wright, R. A. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge: Cambridge University Press.

Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics, 47*, 18–35.

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control, 15*, 118–140.

Zsiga, E. (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics, 28*, 69–102.

Zsiga, E. (2003). Articulatory timing in a second language: Evidence from Russian and English. *Studies in Second Language Acquisition, 25*, 399–432.

Please cite this article in press as: Pouplier, M., et al. Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.01.006