

Purdue University
Purdue e-Pubs

School of Languages and Cultures Faculty
Publications

School of Languages and Cultures

3-22-2017

The Role of Perception in the Typology of Geminate Consonants: Effects of Manner of Articulation, Segmental Environment, Position, and Stress

Olga Dmitrieva
Purdue University, odmitrie@purdue.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/lcpubs>

Recommended Citation

Dmitrieva, Olga, "The Role of Perception in the Typology of Geminate Consonants: Effects of Manner of Articulation, Segmental Environment, Position, and Stress" (2017). *School of Languages and Cultures Faculty Publications*. Paper 17.
<https://docs.lib.purdue.edu/lcpubs/17>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.
Please contact epubs@purdue.edu for additional information.

The role of perception in the typology of geminate consonants: Effects of manner of articulation, segmental environment, position, and stress.

Olga Dmitrieva

School of Languages and Cultures & Linguistics Program, Purdue University, USA

Abstract

The present study seeks to answer the question whether consonant duration is perceived differently across consonants of different manners of articulation and in different contextual environments and whether such differences may be related to the typology of geminates. The results of the crosslinguistic identification experiment suggest higher perceptual acuity in labeling short and long consonants in sonorants than in obstruents. Duration categories were also more consistently and clearly labelled in the intervocalic

Corresponding author:

Olga Dmitrieva, Purdue University, 640 Oval Drive, Stanley Coulter Hall, West Lafayette, IN 47907, USA.

Email: odmitrie@purdue.edu

than in the preconsonantal environment, in the word-initial than in the word-final position, and after stressed vowels than between unstressed vowels. These perceptual asymmetries are in line with some typological tendencies, such as the crosslinguistic preference for intervocalic and post-stress geminates, but contradict other proposed crosslinguistic patterns, such as the preference for obstruent geminates and the abundance of word-final geminates.

Keywords

Geminate typology, perception of consonant duration, Russian, Italian, English

Introduction

One of the predictions of the perceptually-based theories of phonology is that more perceptually distinct pairs of sounds should be used more often as phonemes than sounds which are less distinct. For example, languages appear to prefer vowel systems that are well dispersed in the acoustic vowel space while shying away from skewed inventories

where the available acoustic space is used only partially (Liljencrants & Lindblom, 1972; Lindblom, 1986; Schwartz, Boe, Vallee, & Abry, 1997a, 1997b; Flemming, 2004).

At the same time, it is well known that the acoustic make-up of speech sounds varies considerably across contextual environments, with certain acoustic correlates becoming more or less perceptually prominent or disappearing completely depending on the context (Wright, 2004). For example, vowel formant transitions, which serve as a major cue to the place of articulation in consonants (Delattre, Liberman, & Cooper, 1955), are not available in consonantal environment. It is likely not a coincidence that across languages place contrasts are more common in vocalic than in preconsonantal environments (Fujimura, Macchi, & Streeter, 1978; Ohala, 1990, Hura, Lindblom, & Diehl, 1992; Beddor & Evans-Romaine, 1992; Steriade, 1997, 2001). Such distribution and inventory asymmetries suggest that perception participates in regulating the phonological inventories and the contextual distribution of phonological contrasts.

The present study investigates whether perception could also play a role in shaping the inventories and the contextual restrictions of consonant duration contrasts. As a way to address this question, the study compares the experimentally-determined patterns in the perception of consonant duration against the typological tendencies. The remainder of this section will review the reported typological tendencies in the domains of geminate

inventories and contextual restrictions on geminate distribution. The possible perceptual bases for the observed typological patterns will also be introduced and discussed in this section. The hypothesized perceptual bases provide the direction for the predictions tested in the present experiment.

Geminate inventories: Sonorants versus obstruents

The typological investigations of geminate inventories report patterns that appear amenable to the perceptual explanation. For example, typological studies report that sonorant geminates are less common crosslinguistically than obstruent geminates (Jaeger, 1978; Taylor, 1985; Thurgood, 1993).¹ The comparative rarity of sonorant geminates is hypothesized to stem from a greater acoustic similarity between sonorants and surrounding vowels, especially in terms of intensity, and the resulting difficulty in identifying the boundaries between sonorants and adjacent vowels and estimating sonorants' duration (Podesva, 2000, 2002). This hypothesis was supported experimentally by Kawahara (2007), Kawahara and Pangilinan (in press), and Hansen (2012) who showed that Arabic, English, and Persian listeners were better at identifying and discriminating short and long obstruents than short and long sonorants.

Both the proposed perceptual explanation and the available experimental results point to the idea that the perceptibility of duration in consonants should deteriorate with greater sonority, or intensity, of the consonant. To address this hypothesis in a greater variety of languages the present study compares perception of duration for consonants of different manners of articulation (and sonority levels) by American, Russian, and Italian listeners.

The three language backgrounds of the listeners were chosen due to the range of options they represent with respect to the phonological role of consonant duration. Italian is a classic example of a language with phonemic consonant length. Geminate consonants are found predominantly intervocalically and, in limited cases, before liquids, as in *applicato* ‘applied’ or *soffrire* ‘to suffer’. In Russian, consonant duration is arguably used as a quasi-phonemic, or a ‘facultative’ distinction (Ardentov, 1979). Most Russian geminate consonants are concatenated, i.e. occurring at the boundary of morphemes, although some morpheme-internal gemination, typically in loan-words and words with historically obliterated morphological boundary, is also attested (Ardentov, 1979; Avanesov, 1984; Kolesnikov, 1995; Matusевич, 1976; Panov, 1979). Geminate consonants in Russian are found in a variety of contextual environments, including preconsonantal (e.g. /isslʲedovanije/ ‘research’), word-final (e.g., /stress/ ‘anxiety, stress’), and word-initial (e.g. /ssuda/ ‘loan’), although variable degemination in informal speech is widely attested (Kasatkin & Choj, 1999; Dmitrieva,

forthcoming). Russian and Italian geminate inventories include both sonorants and obstruents (e.g. Russian /vanna/ ‘bathtub’, /gruppa/ ‘group’). In American English, consonant duration is not used phonemically, although phonetic lengthening at the junction of identical consonants is observed with suffixation and compounding (Kotzor, Molineaux, Banks, & Lahiri, 2016).

Prosodic position: Presence or absence of stress on the preceding vowel

Another typological tendency noted by several researchers (Thurgood, 1993; Blevins, 2004) is the correlation between stress and geminates, in particular the fact that geminates tend to occur after stressed vowels. Thurgood (1993) and Blevins (2004) propose that this correlation is a result of the phonologization of the durational increase in the segments which belong to stressed syllables (Cho & Keating, 2009; Turk, 1992). However, it is also possible that perceptual factors are at work here. Podesva (2002), relying on the work by Kato, Tsuzaki, and Sagisaka (1997) suggests that greater differences in intensity between adjacent segments may facilitate perception of duration. Durational distinctions may be easier to detect in obstruents than in sonorants because obstruents are more different from surrounding vowels in terms of intensity. Applying the same reasoning, it can be argued that stressed vowels, being generally louder than unstressed ones, provide a better perceptual environment for the duration-based contrasts in consonants. In addition, due to

their acoustic prominence, stressed vowels may attract greater attention thus providing favorable conditions for discrimination and ultimate preservation of all adjacent contrasts, including durational ones. This hypothesis predicts a perceptual advantage for consonant duration in all stress-adjacent positions, both with preceding and following stress, while a consonant between two unstressed vowels should be in a perceptually disadvantaged position. To test this prediction, the three stress-related contexts are compared in the present study with respect to the perception of consonant duration.

Segmental environment: Vocalic versus consonantal

Scholars also agree that the intervocalic position is where singleton-geminate contrasts are by far the most frequent across geminating languages (Taylor, 1985; Thurgood, 1993; Kraehenmann, 2001; Muller, 2001; Pajak, 2010). Non-intervocalic geminates, on the other hand, are often targeted by neutralization and environment repairs. Cases of surface neutralization of non-intervocalic geminates are attested in Hungarian, Maltese, and Persian, as well as dialects of Arabic and German (Pycha, 2010; Borg, 1997; Mahootian, 1997; Cowell, 1964; Erwin, 1963; Seiler, 2009). Vowel epenthesis occurs before initial geminates in Maltese (Borg, 1997; Hume & Johnson, 2001 and references therein) and Marshallese (Harrison, 1995). Epenthetic schwa after final geminates is reported for Wolof

(Bell, 2003). Similarly, epenthetic vowels can be introduced word-medially between geminates and adjacent consonants, as for example in Hungarian (Ringen & Vago, 2011).

Non-intervocalic geminates can also be subject to additional contextual restrictions. For example, languages such as Finnish and Italian allow consonant-adjacent geminates only if the neighboring consonant is a sonorant.

These tendencies are mirrored in the distribution of another duration-based contrast, the tap-trill distinction, which, as shown by Bradley (2001), is largely restricted to the intervocalic position. Padgett (2003) argues, following Bradley (2001), that taps are more perceptually distinct from trills in the intervocalic environment. Padgett (2003) suggests that the reason for this perceptual advantage lies in the relative ease with which the boundaries of the consonantal segments can be identified in the intervocalic environment. Since perception of duration depends on the reliable detection of the beginning and end points of the segment, this advantage is likely to play a significant role in determining the perceptibility of durational categories.

The prediction that emerges from this overview is that perception of duration should be facilitated by the intervocalic environment and impeded by adjacent consonantal sounds.

The present study tests this prediction by examining perception of duration in both the intervocalic and the consonant-adjacent environment.

Edge geminates: Word-initial versus word-final

There is less agreement in the typological literature regarding the distribution of durational contrasts in non-intervocalic environments. While it is clear that intervocalic position is preferred over word-initial, and word-final ones, it is not obvious whether there is any order of preference between the latter. Related research on the typology of rhotic contrasts (Bradley, 2001) suggests that word-initial position may be more suitable for duration contrasts than word-final and consonant-adjacent ones. Bradley (2001) established the following implicational hierarchy for rhotic contrasts: intervocalic < word-initial < elsewhere (consonant-adjacent, word-final), where a contrast in a given position entails a contrast in the position to the left. Moreover, psycholinguistic research shows that word onsets are of particular importance for lexical access and word recognition and as such are subject to less acoustic distortion in production (for a review see Hawkins & Cutler, 1988). Thus, all contrasts, including durational ones, may be more perceptible in the word-initial position, whether as a result of greater selective attention or due to the high precision of the acoustic realization.

In the typology of geminates, the hypothesis that word-initial contrasts may have a special status finds support in the fact that there are languages where only word-initial geminates are permitted, such as Pattani Malay, Yapese, and Sa'aban (Muller, 2001). However, outside of this small group of languages, initial geminates appear to be quite rare (see Muller, 2001 for an extensive survey) while word-final geminate contrasts are readily found in many familiar languages, such as Hungarian, dialects of German, Arabic, and Norwegian. The general consensus in the literature appears to be that word-initial geminates are less common than word-final one, although a comprehensive survey is yet to be undertaken (Spaelti, 1994; Kraehenmann, 2001, 2011; Thurgood, 1993). The perceptual explanation for this preference, if any, is unclear.

The present experiment tests for the perceptual differences between word-initial and word-final consonant duration. Based on relevant psycholinguistic findings and research on rhotic contrasts, an advantage for the word-initial position is predicted. If such an advantage is found, the apparent typological preference for word-final geminates should be explained by factors other than perception.

Current study

Typological literature reports that geminate inventories are skewed towards obstruents, not sonorants, and that geminates are more readily found in the intervocalic environments with preceding stress, and in the word-final rather than word-initial position. Possible perceptual explanations for some of these tendencies were outlined in the previous research (Bradley, 2001; Padgett, 2003, Podesva, 2002) but not always augmented with experimental evidence (although see Hansen, 2012; Kawahara, 2007; Pajak, 2013). The present work was undertaken to test the perceptual hypotheses experimentally. Specifically, the goal of the study was to determine whether perception of duration was affected by consonants' manner of articulation and contextual environment. The second goal was to verify whether the observed perceptual advantages were in the same direction as the reported typological preferences.

To this end, the study examined listeners' ability to identify the duration category of the consonant (short or long) (1) in consonants of five different manners of articulation; (2) in preceding stress versus following stress versus no stress condition; (3) in intervocalic versus pre-consonantal environments; and (4) in word-final versus word-initial positions. If perception has a connection to the typological tendencies, identification of duration would be more successful in the typologically preferred conditions, i.e. in obstruents, in

consonants located in the intervocalic and preceding stress environments, and in word-final consonants.

To evaluate the relative ‘success’ of identification, properties of the identification function were compared across conditions. Specifically, the steepness of the identification function is considered as an indicator of the *consistency* of categorization and the *precision* of the perceptual boundary between short and long consonants (Aliaga-García & Mora, 2009; McCarthy, Mahon, Rosen, & Evans, 2014; Hutchins, Gosselin, & Peretz, 2010). A steeper identification curve indicates a higher degree of consistency in labelling duration and a clearly defined perceptual boundary. A shallower, more gradient identification slope shows less consistency in categorization and a fuzzier perception boundary. (Note that the research question is not whether ‘true’ categorical perception took place; to establish this it would be necessary to show that discrimination of stimuli was more successful across the perceptual boundary.)

Methods

Stimuli

Stimuli for the perceptual experiment were non-words, pronounced and recorded by the author (a native speaker of Russian).

Stress and Manner of articulation factors: Obstruents versus sonorants in the preceding stress, following stress, and no stress condition. Five minimal triplets (15 words) were created to examine the effects of stress and manner of articulation on the perception of duration. The target consonant (C) was embedded in the stimuli of the structure [koCapu]. Across the members of each triplet, either the first, second, or third vowel of the word was stressed such that the target consonant was preceded by the stressed vowel, followed by the stressed vowel, or not adjacent to the stressed vowel: ['koCapu] – [ko'Capu] – [koCa'pu].² Five different manners of articulation were used as targets (in the order of increasing sonority): a voiceless stop [t], a voiced stop [d], a voiceless fricative [s], a nasal [n], or a liquid [l]: [kotapu], [kodapu], [kosapu], [konapu], and [kolapu].

The Segmental Environment factor: Intervocalic versus pre-consonantal conditions. The stimuli included two minimal pairs (four words), in which the target consonant, in this case always a voiceless coronal fricative [s], was intervocalic or pre-consonantal: [isek] – [islek]; [bisik] – [bismik]. Stress was always final in these stimuli.

Position factor: Word-initial versus word-final conditions. The stimuli consisted of two minimal pairs (four two-word combinations) contrasting the word-final position of the target consonant (a voiceless coronal fricative [s]) with the word-initial position: [pos avap] – [po savap]; [dis ipa] – [di sipa]. In disyllabic words stress was always word-final, thus the

target fricative was always preceded by a stressed vowel and followed by an unstressed vowel. A short pause was inserted between the two words in each stimulus. Pause duration correlated with the number of phones in the stimulus: a shorter pause was introduced in the [pos avap] – [po savap] stimulus (circa 240 ms) than in the [dis ipa] – [di sipa] stimulus (circa 360 ms). This was motivated by the assumption that the compression effect in words with greater number of segments extends to pause duration.

Stimuli were initially recorded with a geminate (long) target consonant and subsequently manipulated in Praat (Boersma & Weenink, 2010) such that 19 versions of each word were created. In each version, the target consonant was of a different duration, starting from the shortest (50 ms) and ascending to the longest (410 ms) in steps of 20 ms. The end-points of the durational continuum were chosen with the goal of incorporating the naturally occurring durational ranges for all the stimuli types as well as keeping both the range and the number of steps identical across conditions. Given that some of the target consonants are naturally short (e.g. singleton liquids and nasals), while others can be considerably longer (e.g. word-final voiceless fricative geminates), the resulting durational range is relatively wide.

The manipulation resulted in the total of 437 experimental stimuli: $(15+4+4)*19$. Stimuli were recorded in a sound-attenuated booth using professional recording equipment. Stimuli

with manipulated duration were resynthesized using the Pitch-Synchronous Overlap and Add (PSOLA) method. The end-points of the target consonants in the original recording were recorded in a Praat text-grid and the duration portion between the end points was manipulated without extracting the target consonant from the surrounding context. All other acoustic parameters of the recorded stimuli were left intact to preserve every possible secondary cue to the duration differences.

Among the possible secondary cues to consonant duration differences is the duration and other acoustic properties of the preceding vowel. For example, in languages, such as Italian, vowels are consistently shorter before geminate than before singleton consonants (Esposito & Di Benedetto, 1999; Rogers & D'Arcangeli, 2004). Fundamental frequency and intensity of the vowel could also be involved in cuing consonant duration (Abramson, 1991, 1999).

With respect to the specific languages of interest in the present study, only preceding vowel duration has been shown to affect perception of gemination in Italian (Esposito & Di Benedetto, 1999; Pickett, Blumstein, & Burton, 1999). Secondary cues to length contrasts in Russian have not been addressed in published literature known to the author. Nevertheless, it is possible that certain acoustic properties of the stimuli, if asymmetrically distributed across the experimental conditions, could act as a secondary cue to gemination

affecting participants' performance in some conditions but not in others. To address this possibility, duration, fundamental frequency, and intensity of vowels, as well as intensity of frication noise and stop bursts were examined across conditions in the naturally recorded experimental stimuli.

Both vowels preceding and following the target consonant were considered. Results showed little variation in the acoustic parameters except in cases where such a variation is expected. For example, vowel duration varied across the three stress conditions: when stressed, the vowel, whether preceding or following, had a longer duration and a higher fundamental frequency and intensity (stressed vowel: 110 ms, 221.5 Hz, 74 dB; unstressed vowel: 58 ms, 202 Hz, 68 dB; averaged across stimuli with targets of different manners of articulation). Vowels also tended to be longer before voiced consonants (/l/, /n/, and /d/) than before voiceless ones (84 ms versus 69.5 ms respectively, averaged across stimuli with different stress placement). Vowels preceding a word-initial target were longer than those preceding a word-final target due the difference in syllable structure (202 ms versus 123 ms, averaged across two stimuli with word-initial and two stimuli with word-final targets respectively): The former belonged to an open syllable (e.g. *possavap*) while the latter was in a closed syllable (e.g. *possavap*).

Intensity of the frication noise for word-initial and word-final /s/ was also examined, demonstrating that word-initial fricatives was higher in intensity than word-final fricatives (63 dB versus 59 dB, respectively, averaged across two word-initial and two word-final stimuli). Intensity range was also higher in word-initial fricatives, suggesting a steeper intensity rise (14 dB versus 7 dB for word-final fricatives). Intensity measurements were performed on the fricatives with the shortest duration of 50 ms.

Finally, burst intensity was examined for stop targets. The comparison between voiced and voiceless stops, averaged across stimuli with different stress placement, showed a difference in terms of burst intensity (67 dB for voiced stops versus 67.5 dB for voiceless stops). Stress placement had an even greater effect on burst intensity. Stops followed by stressed vowels had a greater burst intensity (70 dB, averaged across voiced and voiceless stops) than stops between two unstressed vowels (68.5 dB) or stops preceded by a stressed vowel (63 dB). Burst intensity measurements were performed on stops with an intermediate duration of 210 ms.

Procedure

A forced-choice identification task, implemented in E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA), was used in the experiment. Stimuli were presented to the

participants via high-quality headphones at a comfortable listening level (volume control was indicated to participants). Auditory presentation of the stimuli was accompanied by the orthographic presentation on the computer screen (Courier New font, 32 pt, black on white) in the alphabet of the participant. In the orthographic presentation, target consonants were substituted with a question mark: *ko?apu*.

Participants were instructed to listen to each word and answer as quickly as possible whether the target consonant was short or long by pressing one of the two buttons on the button box or the keyboard (depending on testing location). The corresponding buttons were labeled 'short' and 'long' in the native language of the participant.

Each trial began with the orthographic version of the stimulus appearing on the screen followed in 0.5 sec by the audio presentation. During the 2.5 sec interval after the onset of the audio presentation of the stimulus, participants were expected to provide a response (none of the stimuli exceeded 1.4 sec in duration). If a response was received, the experiment proceeded to the next stimulus after a 2 sec interval of silence and blank screen (duration of the ISI interval was increased to 2.5 sec for Italian participants). If no response was entered, the experiment proceeded to the next stimulus once 2.5 sec elapsed. Thus, each word remained on the screen for the maximum of 3/3.5 seconds.

Stimuli were randomized for each participant and presented in three blocks. The blocks divided the list of unique stimuli into three portions, no stimuli repetitions occurred for the same participant. Participants had an opportunity to take a five-minute break after each block. Each participant began the experiment with a short 10-item practice session. Target consonants in the practice stimuli were selected from the opposite ends of the duration scale to facilitate familiarization with the task. The experiment took approximately 40 minutes to complete.

Participants

Eighty eight (88) listeners participated in the experiment. None reported a speech or hearing disorder. Twenty five (25) of the participants were native speakers of Russian, recruited on the campus of Stanford University and in the neighboring communities. The majority of Russian participants were graduate students at Stanford University and their spouses, between 20 and 30 years of age. As Russian participants were recruited in the United States they all had a significant degree of exposure to English, although their English proficiency varied. Two of the Russian participants were short-term visitors, who resided permanently in Germany and were highly proficient speakers of German (but not English). However, since neither English nor standard German use consonant duration

phonemically, this linguistic background is expected to have little effect on the outcome of the present experiment.

Thirty one (31) of the participants were native speakers of American English. These participants were recruited from the subject pool of the Department of Linguistics of Stanford University. They were students enrolled in the undergraduate linguistics classes, mostly young adults in their early 20s, and were given course credit for their participation. The data reported here were collected from those of American participants who reported no significant exposure to foreign languages. Russian and American participants were tested in the Linguistics Laboratory of Stanford University, using a desktop computer, high quality headphones, and a response box.

The remaining thirty two (32) participants were native speakers of Italian. Italian participants were recruited through word of mouth and fliers at Stanford University and Purdue University (eight participants), and in the communities of the town of Varazze, Italy (24 participants). Italian participants were young adults, mostly in their early 30s. The majority of Italian participants in the present study were natives of Liguria, a region of Italy situated in the north-west of the country, on the coast of the Ligurian Sea, and bordering with France in the west. They were speakers of the local variety of Standard Italian and were not active users of regional dialects.³ Although a fairly consistent exposure to local

dialects can be assumed for all Italian participants, this particular generation of speakers typically developed only a passive knowledge of the dialect. Thus, apart from varied degrees of comprehension proficiency in local dialects, Italian participants recruited in Italy were monolingual speakers of Standard Italian and did not have any prior significant exposure to other languages.

The majority of Italian participants (24) were tested in a quiet room of a private residence in Varazze, Italy, using a laptop computer (IBM Lenovo ThinkPad T60), Sony MDR-NC7 noise canceling headphones, and a keyboard as an input device. Three (3) participants were tested in the Speech and Hearing clinic of Purdue University using the same equipment. Five (5) participants were tested in the Linguistics laboratory of Stanford University using a desktop computer and a response box.

Analysis

The results of the identification experiment were analyzed using a series of Mixed Logistic Models (binary logistic regressions) implemented in IBM SPSS 22 (IBM Corp., Armonk, NY). The omnibus model included Context, Language, Manner (factors), and Duration (covariate) as fixed effects. In the omnibus model, Context was a 7-level aggregate variable which included all levels of the Position, Environment, and Stress factors: word-initial and

word-final; word-medial intervocalic and word-medial preconsonantal (disyllabic words); word-medial with preceding stress, word-medial with following stress, and word-medial not adjacent to stress (trisyllabic words). The omnibus model was supplemented by three follow-up analyses, aimed at providing more details for specific contextual comparisons. Because of the nature of the contextual environments under investigations (for example, a target consonant cannot be word-initial and medial intervocalic at the same time) and to keep the number of experimental items manageable for the participants, not all factors in the design were fully crossed. Specifically, Position and Environment factors were not crossed with Stress and Manner factors or with each other. Therefore, the effects of contextual factors and their interactions with Duration, are better understood within the subsets of data designed to test a specific comparison (e.g., the comparison between word-initial and word-final consonants). To this end, the omnibus model was followed by three additional Mixed Logistic Models, each one dedicated to a specific contextual factor (Environment model and Position model), or a combination of fully-crossed factors, in the case of Stress and Manner variables (Stress/Manner model). Manner was a factor only in the Stress/Manner follow-up model, since all other stimuli were restricted to /s/ targets. In all other respects, the follow-up models were identical to the omnibus model (e.g., all models included Language and Duration as fixed factors). Only the interaction results from

the follow-up models are discussed; full results of the follow-up models are provided in Appendix A.

The results of these statistical analyses are described in detail in the following section. Duration was treated as a continuous variable (a covariate) in all models. Duration was centered by subtracting the mean from each value, such that the new mean was equal to zero. Centering continuous predictors in regression analysis increases interpretability of the regression coefficients and is particularly recommended when the predictor variable does not contain a meaningful zero value (Dalal & Zickar, 2012).

Every two-way interaction between categorical predictors and Duration was tested in all reported models, as these interactions were of the primary theoretical interest in the study.⁴ A significant interaction between Duration and a given contextual factor indicates that the effect of Duration on the categorization of the stimuli was not consistent across the levels of the contextual predictor. That is, as target consonant becomes longer, the rise in the frequency of ‘long’ categorization is more pronounced at one level of the contextual factor (e.g. word-initial targets) than another (e.g., word-final targets). A more pronounced rise indicates a steeper identification function, suggesting a more consistent labelling of consonant duration and a more clearly defined perceptual boundary. These statistical

indications were confirmed via examining the averaged identification curves across the relevant contextual conditions.

A random intercept for subject, and a random intercept for subject by item were significant according to the Wald statistic ($p < 0.001$) in all models and were included in the models reported below (variance component covariance structure). Robust estimation was used for the tests of fixed effects and coefficients (robust covariances) in all models.

Results

Figure 1a-c demonstrates the effect of target duration on duration categorization across stress, environment, and position conditions. Figure 1a shows that the two categorization curves for targets followed by stress and not adjacent to stress overlap almost perfectly, while the curve for targets preceded by stress stands out with a marginally steeper slope, shorter perceptual boundary (around 140 ms), and an overall greater proportion of ‘long’ responses, especially at the higher end of the duration scale. Figure 1b demonstrates that the categorization curve for intervocalic targets is steeper in the middle portion and flatter in the tails than the curve for preconsonantal targets. Figure 1c shows that the categorization curve for initial targets rises more steeply than the curve for final targets, reaching higher percentages of ‘long’ categorization earlier than the final curve.

<Insert Figure 1a-c about here>

The subtle but discernable differences in the steepness of the identification slopes, combined with statistical results presented below, suggest that labelling of duration was more consistent in the preceding stress, intervocalic, and word-initial contexts, compared to the following and non-adjacent stress, preconsonantal, and word-final contexts. Due to the steeper identification functions, perceptual boundary is more clearly defined in the former contexts, suggesting more refined durational categories.

The omnibus statistical model included four fixed effects: Context, Language, Manner, and Duration. The dependent variable was the binary response: ‘short’ or ‘long’. Degrees of freedom varied across tests (Satterthwaite approximation) for the omnibus model only. The model achieved 92% classification accuracy, which was 40% better than the by-chance accuracy rate. The significant main effects and interactions in the omnibus model are summarized in Table 1.

Table 1: Results of the omnibus Mixed Logistic Regression with binary duration categorization response as the dependent variable.

<i>Factor</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Context	40.475	6	1,338	.000
Manner	39.074	4	2,607	.000

Language	8.777	2	215	.000
Duration	246.582	1	36,295	.000
Duration*Context	7.883	6	36,295	.000
Duration*Manner	17.496	4	36,295	.000
Language*Language	5.005	2	36,295	.007

The details for the terms with significant coefficients are provided in Table 2 and discussed below.

Table 2: Terms with significant coefficients in the omnibus Mixed Logistic Regression with binary duration categorization as the dependent variable. Reference category: ‘short’.

	<i>Model Term</i>	<i>Coeff-ent</i>	<i>Std. Error</i>	<i>p</i>	<i>Exponented Coeff-ent</i>
	Intercept	-1.119	0.222	.000	0.326
Reference: <i>Initial</i>	Elsewhere stress	2.210	0.312	.000	9.115
	Preceding stress	2.754	0.285	.000	15.709
	Following stress	2.214	0.294	.000	9.154
	Preconsonantal	0.799	0.209	.000	2.224
	Intervocalic	1.048	0.264	.000	2.853
	Final	-0.543	0.164	.001	0.581
Reference: /s/	/l/	1.871	0.249	.000	6.496
	/n/	1.782	0.284	.000	5.944
	/t/	-1.576	0.216	.000	0.207
Reference: <i>English</i>	Russian	1.008	0.246	.001	2.740
	Italian	0.881	0.246	.000	2.412
Duration	Duration	0.023	0.002	.000	1.024

Reference: <i>Duration*Initial</i>	Duration*Elsewhere stress	0.008	0.002	.000	1.008
	Duration*Preceding stress	0.010	0.002	.000	1.010
	Duration*Following stress	0.009	0.002	.000	1.009
	Duration*Preconsonantal	0.004	0.002	.023	1.004
	Duration*Intervocalic	0.007	0.002	.000	1.007
	Duration*Final	-0.003	0.001	.056	0.997
Reference: <i>Duration*/s/</i>	Duration*/d/	-0.010	0.002	.000	0.990
	Duration*/t/	-0.016	0.002	.000	0.984
Reference: <i>Duration*English</i>	Duration*Russian	0.010	0.003	.002	1.010

The Context factor. The positive and significant coefficients for all but one level of the Context factor indicate that in comparison to the reference level (word-initial context), almost all other types of context *increased* the likelihood of ‘long’ categorization. For example, the odds of medial targets preceded by stress to be categorized as ‘long’ were as much as 15 times higher than those of word-initial targets: $\exp(2.754)=15.7$.

Word-final targets were the only ones which were *less likely* to be categorized as ‘long’ than the reference word-initial targets. The exponentiated value of 0.581 of the marginally-significant negative coefficient for final targets indicates that the odds of word-final /s/ to be categorized as ‘long’ were about 42% lower than those of word-initial /s/.

These results suggest that consonants on the edges of words, especially the word-final ones, are at a disadvantage with respect to the perception of length, when compared to

word-medial consonants, across a used scale of duration. However, this result should be interpreted with caution, since in this comparison other factors are conflated with context and could confound the results. In particular, edge targets were restricted to voiceless coronal fricatives /s/, while word-medial ones included targets of other manners of articulation, such as sonorants and stops. In addition, stimuli incorporating edge targets were monosyllables, while stimuli incorporating medial targets were di- and tri-syllables.

To verify the results of the omnibus model with respect to the effects of contextual factors on the categorization of duration, three follow-up mixed effects logistic regressions were conducted, each one restricted to the set of stimuli designed to make a specific contextual comparison: word-initial versus word-final, intervocalic versus preconsonantal, preceding stress versus following stress versus no adjacent stress. In all three models the effect of Context factor was significant, with positive coefficients indicating that the odds of ‘long’ categorization were significantly higher in the word-initial than word-final position ($\beta=0.591$, $e^\beta=1.805$, $SE=0.176$, $p<0.01$), in the intervocalic than preconsonantal environment ($\beta=0.274$, $e^\beta=1.315$, $SE=0.135$, $p<0.05$), and in the preceding stress than in the elsewhere stress condition ($\beta=0.543$, $e^\beta=1.721$, $SE=0.142$, $p<0.001$).

The Manner factor. The positive and significant coefficients for /n/ and /l/ in the omnibus model indicate that these targets were more likely to be categorized as ‘long’ in comparison to /s/, the reference category. The exponentiated values show that the odds of ‘long’ categorization were about 6 times higher for sonorant targets than for /s/. A negative and significant coefficient for /t/ indicates that /t/-targets were less likely than /s/ to be categorized as ‘long’. The exponentiated value demonstrates that the odds decrease by 0.207 or about 80% for /t/ compared to /s/.

Thus, all else being equal, sonorant targets are more likely than /s/ to be perceived as ‘long’, while obstruent targets, /t/ in particular, are less likely than /s/ to be perceived as ‘long’, across the used duration scale. However, it should be noted that /s/-category in the omnibus analysis included word-edge /s/ as well as /s/ in the di- and trisyllabic stimuli, while stimuli representing other manners were limited to tri-syllables. For a more controlled comparison between manners of articulation, results of the follow-up model restricted to the stimuli designed to compare consonants of different manners of articulation (incorporated in a factorial design into the three stress conditions) were examined. In the follow-up model, the effect of Manner was significant, with significant positive coefficients indicating that the likelihood of ‘long’ categorization was higher for sonorants /l/ and /n/ than for /s/, the reference category ($\beta=1.789$, $e^\beta=5.985$, $SE=0.253$, $p<$

0.001 and $\beta=1.782$, $e^{\beta}=5.944$, $SE=0.285$, $p<0.001$, respectively). A significant negative coefficient for /t/ indicated that for this target the likelihood of ‘long’ categorization was lower than for /s/ ($\beta=-1.572$, $e^{\beta}=0.208$, $SE=0.215$, $p<0.001$).

The Language factor. The positive and significant coefficients of 1.008 for Russian and 0.881 for Italian indicate that in comparison to English (reference category), Russian and Italian participants were more likely to categorize target sounds as ‘long’ across contexts and manners. The exponentiated values of 2.740 for Russian and 2.412 for Italian participants indicate that the odds of Russian participants to categorize target sounds as ‘long’ were almost 3 times greater than those of English-speaking participants; while the odds of Italian participants to categorize target sounds as ‘long’ were approximately two times greater than those of English-speaking participants.

The effect of Language was also significant in the Stress/Manner follow-up model, and marginally significant in the Environment follow-up model. The direction of the effect was the same as in the omnibus model: Russian or both Russian and Italian groups of participants were associated with a higher likelihood of ‘long’ categorization.

The Duration factor and its interaction with categorical predictors. A positive and significant coefficient for the Duration variable (0.023) in the omnibus model suggests that a longer duration of the target consonant increased the odds of ‘long’ categorization. However, the presence of interactions with categorical variables in the model means that this coefficient is valid only for certain levels of categorical predictors, specifically for *initial /s/ targets* categorized by *English-speaking participants* (the baseline level). The exponentiated value of the coefficient (1.024) shows that, under these conditions, the odds of ‘long’ categorization increased by 2.4% for every ms of increase in target’s duration. However, the fact that the interactions between Duration and all three categorical predictors (Context, Manner, and Language) were significant indicates that the effect of duration was not consistent across the levels of these categorical variables.

For example, *the interaction between Context and Duration* was significant with positive coefficients at all levels of the Context factor, except final. This indicates that the baseline positive effect of duration on the odds of ‘long’ categorization is further magnified by all contexts, except final. For example, for targets preceded by stressed vowels the increase in the odds of ‘long’ categorization is $\exp(0.023+0.01)=1.033$ or 3.3% for every ms of increase in target’s duration, compared to 2.4% of the baseline increase. In contrast, in the final context, the odds of ‘long’ categorization increase by $\exp(0.023-0.003)=1.02$ or only

2% with every ms of increase in target's duration, compared to 2.4% of the baseline increase. Thus, word-final context *decreased*, rather than magnified, the baseline effect of duration on the odds of 'long' categorization.

The interaction between Duration and Context was also significant or near-significant in the three follow-up models, positive coefficients demonstrating that preceding stress ($\beta=0.002$, $e^\beta=1.002$, $SE=0.001$, $p<0.01$), intervocalic environment ($\beta=0.004$, $SE=0.002$, $e^\beta=1.004$, $p<0.05$), and initial position ($\beta=0.003$, $e^\beta=1.003$, $SE=0.002$, $p=.073$) *magnified* the positive baseline effect of duration compared to non-stress adjacent condition, preconsonantal environment, and word-final position.

Figure 2 demonstrates the average categorization curves by manner of the target consonant. It shows that the categorization curves for /d/ and /t/ targets are not as steep as the curves for /s/, /n/, and /l/. Besides, the right tails of the /d/ and /t/ curves show more variability in responses than the curves for /s/ and sonorant targets. This suggests that labeling of duration was less consistent for stop consonants than for fricatives and sonorants, resulting in a fuzzier perceptual boundary.

<Insert Figure 2 about here>

In agreement with the tendencies suggested by Figure 2, the interaction between *Duration* and *Manner* was significant in the omnibus model. The significant negative coefficients at the levels of /d/ and /t/ indicate that the baseline positive effect of duration on the odds of ‘long’ categorization is counteracted by the negative effect of stop manner, resulting in the shallower identification curves. The odds of ‘long’ categorization increase only by 0.7% ($\exp(0.023-0.016)=1.007$) and 1.3% ($\exp(0.023-0.010)=1.013$), for /t/ and /d/ respectively, with every ms of increase in target’s duration, compared to 2.4% of the baseline increase.

The interaction between Duration and Manner remained significant in the follow-up model, negative coefficients indicating that stop manner of articulation *diminished* the positive baseline effect of Duration on the likelihood of ‘long’ categorization ($\beta=-0.016$, $e^{\beta}=0.984$, $SE=0.002$, $p<0.001$ and $\beta=-0.009$, $e^{\beta}=0.991$, $SE=0.002$, $p<0.001$, for /t/ and /d/ respectively).

Figure 3 demonstrates the categorization curves for participants with different language backgrounds in the subset of data from the Stress/Manner model. It shows that the identification curve for Russian participants is steeper than the Italian and especial English

one, suggesting greater consistency of duration labelling by Russian participants and resulting in a more clearly defined perceptual boundary.

<Insert Figure 3 about here>

As suggested by the visual examination of the identification curves in Figure 3, *the Duration by Language interaction* was significant in the omnibus model. A positive interaction coefficient at the level of Russian indicates that the baseline effect of duration on the odds of ‘long’ categorization is further *increased* if participants are Russian, as compared to English, resulting in a steeper identification curve. The odds of ‘long’ categorization by Russian-speaking participants increased by 3.3% ($\exp(0.023+0.010)=1.033$) for every ms of increase in target’s duration, compared to 2.4% of the baseline increase.

Duration by Language interaction remained significant in two out of three follow-up models (trending towards significance in the Position model, with $p=0.07$). The direction of the effect consistently showed that Russian participants magnified the effect of Duration on the odds of ‘long’ categorization, compared to the English-speaking participants.

Summary of the results

The results of the omnibus model showed that some contextual environments and types of consonants were associated with a greater likelihood of ‘long’ categorization than others. In particular, consonants in medial, preconsonantal and especially intervocalic environments, won over word-edge consonants (e.g., word-initial ones) in terms of the odds of ‘long’ categorization. The tendency for more frequent ‘long’ categorization was especially pronounced for targets preceded by stressed vowels. Word-final environment, on the other hand, reduced the chance of ‘long’ response, even in comparison with another word-edge position, the word-initial one.

In terms of manner of articulation, sonorant consonants, liquids and nasals, were more likely to be perceived as ‘long’ than the reference category of voiceless coronal fricatives. Voiceless stops, on the contrary, were associated with a decrease in the likelihood of ‘long’ categorization, compared to /s/. Finally, Russian and Italian-speaking participants were more willing to assign a ‘long’ status to the target consonants than English-speaking participants.

Even more revealing than the main effects were the interactions in the omnibus model. In general, unsurprisingly, greater target duration tended to increase the likelihood of ‘long’ categorization. However, significant interactions with all three categorical variables demonstrated that a positive effect of duration could be either magnified or diminished

when combined with certain levels of categorical predictors. Specifically, all word-medial contexts, and especially the preceding stress condition, *magnified* the effect of duration. Targeted comparisons in the follow-up models further clarified that the preceding stress condition magnified the effect of duration compared to the non-adjacent stress condition. The intervocalic environment magnified the effect of duration compared to the preconsonantal environment and the word-initial position tended to magnify the effect of duration compared to the word-final position.

The positive effect of duration was also increased if targets were categorized by Russian participants. The word-final context and stop targets, on the other hand, counteracted the effect of duration, *reducing* its positive effect on the odds of ‘long’ categorization.

Discussion

The significant main effects in the statistical models demonstrated that the likelihood of ‘long’ categorization was greater for certain contextual environments and groups of speakers. Some of these effects are not entirely surprising and can find straightforward explanations in the well-known facts about the nature of languages, contexts, and consonants involved. For example, English speakers’ reluctance to label target consonants as ‘long’ could stem from the absence of phonemic or even semi-phonemic consonantal

length distinction in American English (see, for example, Pickett and Decker (1960), who showed that for English listeners consonant /p/ had to be over 250 ms long to trigger the perception of *top pick* instead of *topic*).

Furthermore, different perceptual boundaries also effect the overall odds of ‘long’ categorization, since a larger proportion of stimuli for a given duration is perceived as long if the perceptual boundary between singletons and geminates is at a lower durational value (e.g. 100 ms instead of 200 ms). Perceptual boundaries, in their turn, tend to correlate with differences in average consonant duration, such that shorter consonants are associated with a boundary at a lower value (Pickett & Decker, 1960). Accordingly, in the present results we observe shorter perceptual boundaries and greater frequencies of ‘long’ responses for shorter sonorant consonants as opposed to longer obstruent consonants (see Figure 2).

Similarly, segments are likely to be shorter in multisyllabic than in monosyllabic words (Baum, 1992; Lehiste, 1972). Besides, word-initial and word-final segments are typically lengthened due to the domain boundary enhancement phenomenon (Oller, 1973; Byrd, 1993; Turk & White, 1999; Turk & Shattuck-Hufnagel, 2000, among others). Thus, a greater frequency of ‘long’ responses for all word-medial targets as compared to the word-initial and word-final ones can, at least in part, be attributed to a shorter perceptual boundary.

However, of the greatest interest in the statistical results were the presence and the nature of the interactions between duration and contextual factors. An interaction which increased the positive effect of duration on the odds of ‘long’ categorization indicated a steeper categorization slope and a clear perceptual boundary between the categories. Such a perceptual advantage was detected for the intervocalic, preceding stress, word-initial, and sonorant consonants and for the Russian group of participants, as compared to the preconsonantal, non-stress adjacent, word-final, and obstruent consonants and the English-speaking group of participants. The implications of these findings and their consistency with the typological tendencies are discussed in the following sections.

Manner of articulation and geminacy

The most controversial result of the present investigation is the listeners’ response to the consonants of different manners of articulation. Existing typological studies suggest that sonorant geminates are less commonly found across languages than obstruent geminates (Jaeger, 1978; Taylor, 1985; Thurgood, 1993). Available experimental results support the hypothesis that this trend may be due to the greater difficulty inherent in identifying duration in sonorants (Hansen, 2012; Kawahara, 2007; Kawahara & Pangilinan, in press). The proposed explanation for this asymmetry is that sonorants, being most vowel-like of the consonants, may be more difficult to perceptually separate from the surrounding

vowels, which complicates the task of determining the sonorant's duration. Thus, as the sonority-wise similarity between the target consonant and surrounding sounds increases, the perceptual difficulty in determining the target consonant's duration should increase as well. Given that the most common contextual environment for length contrasts is vocalic, the sonorant length contrasts are in a disadvantaged position with respect to perceptibility, and thus are expected to be avoided by languages.

However, the results of the current experiment do not agree with previous experimental findings and the proposed typological trend these findings are called upon to explain. The general pattern that emerged is that high-sonority consonants, which are characterized by the sustained acoustic energy throughout their duration, were categorized into short and long in a more consistent fashion than low-sonority consonants. Low-sonority consonants, represented here by voiceless and voiced coronal stops, have little or no acoustic energy for the greatest part of their duration (i.e. a silent closure for voiceless stops, and a low-frequency low-intensity voicing signal for voiced stops). For these, differences in duration were perceived in a more gradient fashion, suggesting an overall lower degree of consistency in judging the duration of low-sonority consonants in terms of binary categories. Thus, listeners were better at defining duration categories for intervocalic *sonorants* (liquids and nasals) than for intervocalic obstruents, voiceless stops in particular,

despite the high degree of acoustic similarity between sonorants and the surrounding vowels. A sibilant fricative [s], which is also characterized by a sustained and salient acoustic energy throughout its duration, tended to pattern with sonorants.

It is possible that preceding vowel characteristics have contributed to the observed effect. Acoustic analysis of the stimuli showed that vowels before sonorants were *longer* than those before voiceless obstruents. While there is no prior evidence to suggest that longer preceding vowels create better perceptual conditions for the categorization of duration, this possibility is suggested by the present pattern of results.

Another possible explanation for the sonorant advantage detected in the present study may be found in the fact that sonorant and fricative consonants are characterized by a continuous acoustic signal. In other words, higher sonority consonants, such as fricatives, liquids, and nasals, may be viewed as *filled* intervals in contrast to lower sonority consonants, such as stops, which may be viewed as *empty* intervals. There is some evidence from the psychoacoustic literature that the duration of filled intervals (typically tones) is perceived more accurately by humans (Rammsayer, 2010) and some non-human species (Santi, Miki, Hornyak, & Eidse, 2005) than the duration of empty intervals (e.g., periods of silence marked by clicks). This asymmetry has been attributed to the possibility that the neural pacemaker mechanism used to judge duration is stimulated more effectively

by continuous physical stimuli. This results in a greater neuron firing rate and a better temporal resolution for the filled intervals than for the empty intervals (Rammsayer, 2010; Wearden, Norton, Martin, & Montford-Bebb, 2007). Moreover, higher intensity of the stimulus may further contribute to the speeding up of the internal pacemaker mechanism (Matthews, Stewart, & Wearden, 2011), suggesting that duration of the high-intensity continuous sounds, such as sonorants, should be judged more accurately than duration of the low-intensity continuous sounds, such as fricatives (e.g., Goedemans (1998) showed that duration differences were better perceived in /m/ than in /s/ for Dutch listeners). This is reminiscent of the pattern of the results observed in the present study.

However, the filled interval advantage appears to be highly dependent on the experimental methodology and the base duration of the intervals (Grondin, 1993; Plourde, Meilleur-Wells, Gamache, Dionne, & Grondin, 2008). Also, it is not known how the effect established for non-speech stimuli translates into natural speech perception. Furthermore, if filled interval advantage is real for natural speech perception, it is not clear why some phonological systems should exclude sonorant but not obstruent length contrasts from their inventories (Kawahara, 2007; Kawahara & Pangilinan, in press; Podesva, 2000, 2002), insofar as such reports are representative of the general trend. The lack of consensus in the

psychoacoustic and linguistic literature on the topic demonstrates that more research on the perception of duration as function of the manner of articulation and sonority is needed.

Stress and geminacy

The investigation of the effect of stress on the perception of duration revealed an effect in the predicted direction: Listeners were more consistent in labelling consonant duration in the preceding stress condition. However, the following stress condition was not statistically different from the non-stress adjacent condition, suggesting that it is not simply adjacency to the stressed vowel but also its location that matters for the perception of duration. Thus, it appears unlikely that the detected advantage for the perception of duration in preceding stress condition stems from the higher intensity of the adjacent stressed vowel. If this was the factor triggering the perceptual advantage, following stress condition would be expected to produce a similar effect, since both preceding and following vowels in the present stimuli demonstrated an increase in intensity when stressed. A vowel duration differences was also present in the stimuli. Stressed vowels preceding targets in the preceding stress condition were longer and louder than unstressed vowels preceding targets in the following stress condition. Although there is no a priori reason to expect this, it is possible that the difference in vowel duration affected the consistency of labeling duration in consonants. A higher burst intensity in stop consonants could also help demarcate the

target consonant and facilitate the estimation of its duration. However, burst intensity was the lowest in the preceding stress condition and thus an unlikely contributor to the more consistent labeling of duration in this environment.

The diverging results for the preceding and following stress conditions may also be related to the asymmetries between the onset and the coda of the syllable in the perception of duration. Goedemans (1998) demonstrated on the example of monosyllabic stimuli that Dutch listeners were better at estimating duration changes in the codas than in the onsets. In the present experiment, longer target consonants preceded by stress were, at least partially (assuming the ambisyllabicity of intervocalic geminates), in the coda of the stressed syllable. Target consonants followed by stress were in the onset of the stressed syllable. Assuming that stress, acting as the attention-getter, provides some perceptual advantage, consonants in the coda of the stressed syllable are then perceived more accurately in terms of duration than those in the onset of the stressed syllable. The exact psychoacoustic or linguistic triggers of this asymmetry are, however, unclear (for relevant discussion see Goedemans, 1998).

These results suggest that the reported typological preference for geminate consonants to occur after stressed vowels (Blevins, 2004; Thurgood, 1993) may be due to the fact that, in perception, duration categories are more clearly defined in the preceding stress condition.

In addition, the connection between stress and geminacy is well established in the phonology of syllable weight (Davis, 1994, 2011; Hayes, 1989). A geminate coda of the stressed syllable renders the syllable ‘heavy’, or bimoraic, in many phonological systems, satisfying a common bimoraic requirement for stressed syllables in weight-sensitive languages. Onsets typically do not contribute to syllable weight. Thus, geminate consonants are particularly common in the codas of stressed syllables for weight-sensitive languages. It is possible that the criteria of the phonological weight system are attuned to the perceptual preferences: Additional duration (gemination) is more perceptible in codas and therefore it is allowed to contribute phonological weight. This argument is advanced in Goedemans (1998) to explain the general crosslinguistic weightlessness of syllable onsets (see also the work by Gordon (2002, 2006) on the relationship between segmental duration and syllable weight).

Segmental environment and geminacy

The results of the identification experiment showed a benefit of the intervocalic environment for duration categorization. Compared to the preconsonantal (pre-sonorant) targets, intervocalic consonants were categorized by participants with a higher degree of consistency, resulting in a more clearly defined perceptual boundary and more refined duration categories. Given such a perceptual asymmetry between the contextual

environments, it is reasonable to expect that intervocalic length contrasts would be more common across languages than consonant-adjacent ones, which is exactly what the typological literature reports (Kraehenmann, 2001; Muller, 2001; Pajak, 2010; Taylor, 1985; Thurgood, 1993).

The trigger of this asymmetry is unclear. The acoustic-similarity hypothesis (Podesva, 2000, 2002) suggests that the duration of voiceless obstruents, such as fricatives, should be easier to estimate in the intervocalic environment thanks to the marked acoustic dissimilarities between obstruents and vowels. However, the results of the manner condition of the present experiment do not support this view (the duration of vowel-like sonorants was judged more consistently than the duration of voiced and voiceless stops in the intervocalic environment). An alternative explanation is that high sonority context alone aids in duration estimation, independently of the type of the target consonant.

In the present experiment only two environments were compared – intervocalic and sonorant (liquid or nasal) adjacent. However, more detailed predictions can be made based on the assumption that the ability to perceive duration of the consonant depends on the degree of sonority of the adjacent sounds. Specifically, as the adjacent sounds become less vowel-like in quality (i.e. obstruents or fricatives instead of sonorants) and therefore lower in sonority, the ability to estimate target's duration should deteriorate even further. This

prediction can be tested in the design where the perception of consonant duration in the intervocalic environment is compared to a number of consonant-adjacent environments, with adjacent consonants varying in the degree of sonority. The experiment is currently in progress and the initial results confirm the hypothesis that listeners are less capable to detect duration differences for obstruents in low-sonority environment than in high-sonority environment (see also a study by Kawahara and Pangilinan (to appear), which tested the effects of the degree of amplitude changes on categorization and discrimination of length contrasts).

Word-edge position and geminacy

The results of the comparison between the word-initial and the word-final consonants showed that listeners were somewhat more consistent in categorizing consonant duration in the word-initial than in the word-final position. Preceding vowel duration is unlikely to have triggered this effect, since an adjacent preceding vowel was available only in the word-final condition (in the word-initial condition, a pause separated preceding vowel from the target consonant). It is possible that greater psychoacoustic salience of the word-initial position created better conditions for perception of duration in the word-initial targets. In addition, the properties of the fricative itself could have contributed to the perceptual difference between the two positions. Acoustic analysis showed that word-initial fricatives

in the stimuli adopted for the experiment were of a greater intensity and had a steeper intensity rise than word-final ones. These intensity differences could have facilitated the identification of duration categories in the word-initial condition.

Thus, word-initial position appears to be more suitable for consonant duration contrasts from the perceptual stand-point, whether due to attentional or acoustic factors. Based on this result, one should expect to find more consonant length contrasts in the word-initial than in the word-final position across languages. This prediction, however, is not supported by the reported typological observation, which appear to be in favor of the word-final gemination contrasts (Taylor, 1985). The juxtaposition of the current experimental results and the typological tendencies suggests that perceptual factors are not the ones driving the phonological typology in the case of word-final geminates.

These results are at odds with Goedemans (1998) results which showed a greater sensitivity to duration changes in monosyllabic codas (equivalent to word-final position) than in monosyllabic onsets (equivalent to word-initial position). A possible explanation lies in the methodological differences. In the present experiment, both the word-final and the word-initial target was preceded or followed by a vowel-flanked word (after a pause). In Goedemans (1998) stimuli were presented in pairs with comparison stimuli, which always followed the reference stimulus. Thus, coda consonants were followed by another stimulus,

which could have facilitated the estimation of their duration, while onset consonants were presented in an absolute initial position, which could have complicated the estimation of their duration.

A possible explanation for the crosslinguistic prevalence of word-final geminates lies in the involvement of geminate consonants in the phonology of syllable weight. Phonological evidence suggests that word-final geminate, but not singleton, codas can make a syllable ‘heavy’ for the purposes of processes such as stress assignment. For example, in many dialects of Arabic, word-final syllables attract stress only if closed by a consonant cluster or a geminate consonant (e.g., Erwin, 1963). Unlike medial codas, word-final singleton codas often do not count for syllable weight, a phenomenon typically referred to as extrametricality (Hayes, 1980; Liberman & Prince, 1977). Thus, a geminate coda of the word-final syllable is often in the unique position to supply the additional weight when it is required by the phonology of the language. These exceptional weight-bearing properties of the coda geminates, word-final geminates in particular, make them indispensable for weight-sensitive languages with word-final primary or secondary stress, while word-initial geminates have not been associated with an equally important role in the phonology of syllable weight (although see Topintzi, 2008 on moraic onset geminates). This might explain the overall prevalence of word-final geminates as compared to the word-initial

geminate, despite the fact that word-final contrasts appear to be perceptually inferior to the word-initial one.

It is worth mentioning, however, that the relatively low perceptibility of duration categories in the word-final position does not go unaddressed in languages with final geminates. Several phonetic phenomena associated with final geminates may be directed at remedying their sub-optimal perceptual qualities. For example, in Wolof, a short schwa-like vowel is introduced after the final geminate consonant, which in effect turns the word-final position into the intervocalic one (Bell, 2003). In Tashlhiyt Berber the ratio between geminate and singleton stops is higher in the word-final than in the intervocalic position, which is likely to enhance the perceptual difference between singletons and geminates in the word-final position (Ridouane, 2007). Some dialects of Arabic neutralize the final length contrast on the surface (i.e. final geminates are pronounced as singletons), while maintaining the weight-bearing properties of the underlying final geminates (Cowell, 1964; Erwin, 1963). When neutralization applies, there is no longer a need for the listeners to discern final geminates from final singletons; instead, the stress placement itself may signal the fact that the final syllable contains an underlying geminate.

It should also be noted that the only types of word-initial and word-final consonants tested in the study were voiceless coronal fricatives. It is known that other types of consonants,

such as voiced and voiceless stops, nasals and liquids are allowed in these positions by a wide variety of languages. As an extreme example, Tashlhiyt Berber (Ridouane, 2007), Thurgovian Swiss German (Kraehenmann 2001), Cypriot Greek (Armosti, 2009; Arvaniti and Tserdanelis, 2000; Muller, 2001), and Pattani Malay (Abramson, 1991, 1999, 2003) allow voiceless stop geminates in the word-initial position. This discrepancy between the restricted experimental conditions and the naturally occurring variety limits the generalizability of the current findings. The voiceless coronal fricatives were chosen for the present experiment with the goal of obtaining an estimation for the behavior of the continuant consonants in the word-initial and word-final environments.

What concerns stop consonants, especially voiceless stops, duration perceptibility in the word-initial position can be expected to be worse than in the word-final position, due to the lack of acoustic cues to the onset of silent consonant closure (to fare better, the word-final stops must be released). Accordingly, secondary cues have been shown to support word-initial durational contrasts in stops in some languages (e.g. intensity and fundamental frequency in Pattani Malay and aspiration in Cypriot Greek). Thus, there are reasons to believe that the results of the current experiment may not generalize to voiceless stops.

The results nevertheless suggest that the crosslinguistic preference for word-final geminates over the word-initial ones is not likely to be due to the overall perceptual

superiority of the former. In fact, the perceptual account would predict a gradual elimination of word-final contrasts from the phonological inventories, at least for continuant consonants. However, the loss of contrast is not the only remedy for the perceptual difficulties. It appears that some languages employ environment repairs and phonetic enhancement to boost the perceptual distinctiveness of final (and initial) duration contrasts. This approach allows languages to retain final geminates in particular for their useful weight-bearing properties without compromising the perceptual efficiency of the system of phonological contrasts.⁴

Language-specific effects

The results of the experiment showed that there was a clear difference in the perceptual behavior of the three participant groups. In particular, Russian and Italian participants consistently outperformed American listeners in terms of the overall odds of ‘long’ response. Russian participants also showed a greater consistency of duration categorization and a ‘sharper’ perceptual distinction between short and long consonants than English-speaking participants. The latter difference may be attributed to the shared-language benefit available to Russian participants. The stimuli for the experiment were recorded by a native speaker of Russian and thus the segments composing the experimental stimuli were represented by the allophones characteristic of Russian or Russian-accented speech. For

example, the high central vowel used in some stimuli is found in the Russian but not Italian or English inventory. This has likely boosted the accuracy of the auditory recognition of the target and non-target sounds by Russian participants. In addition, some of the phonotactic features of the stimuli may have been more familiar to Russian participants. In Russian, long consonants can be found in a great variety of contextual environments, including preconsonantal, word-initial, and word-final, as well as in all three stress-related environments, although they are subject to frequent degemination in casual speech and are rarely meaning-differentiating. Thus, the advantage for the perceptual categorization of duration observed for Russian participants cannot be attributed to the universally greater acuity for the perception of duration in this group of listeners. Rather, the specific characteristics of the stimuli used in the experiment created an advantage in these particular experimental conditions. American participants, on the other hand, categorized duration in a more gradient fashion than Russian listeners. This pattern of responses is compatible with the fact that American English does not use consonant length even in a marginally phonological role. Thus, American participants have little experience in categorizing consonant duration. Nevertheless, by and large, American participants were able to perform the task and had little difficulty understanding the instructions. This suggests that duration

is a particularly salient acoustic property and can be used quite successfully in categorizing speech sounds even by speakers of the languages which do not use duration phonemically.

Finally, Italian participants could be expected to perceive duration differences in the most consistent fashion given that Italian is a language where consonant length is truly phonemic. Being able to identify long and short consonants, therefore, must be an important part of the speech perception skills of the native Italian listeners. In contrast, while geminate consonants are quite common in Russian, their functional load is comparatively low, i.e. few minimal pairs rely on the distinction between short and long consonants. In addition, it is likely that the overall frequency of geminate consonants in Russian language is lower than in Italian. Moreover, unlike Italian geminates, Russian ones are subject to variable degemination in speech.

Despite their extensive experience with phonemic consonant length, Italian participants were not statistically different from English-speaking participants in terms of the consistency of categorization and steepness of the categorization curve. This can be attributed to the fact that Italian participants' perception of duration may be strongly dependent on the way this contrast is implemented in native phonotactics. While speakers of English, and to some degree Russian, start with a 'blank slate', meaning that length contrasts are equally unexpected (or expected) in all contexts, Italian listeners have come to

expect length contrasts in certain environments but not in others. For example, contrastive geminates are not found in Italian word-initially (apart from the phonosyntactically conditioned initial lengthening known as *Raddoppiamento Sintattico*), and consonants in general are rare in the word-final position. Similarly, consonant-adjacent geminates are rare in Italian and have a low functional load. It is possible that Italian participants were less attuned to variation in consonant duration in these environments, which would contribute to their overall less confident performance in the task. Italian listeners are also known to use preceding vowel duration in addition to consonant duration itself in determining the geminate status of the consonant (Esposito & Di Benedetto, 1999; Pickett, Blumstein, & Burton, 1999). Preceding vowel duration was not manipulated in the stimuli. Instead, it was kept constant regardless of the duration of the target consonant. The resulting vowel-to-consonant duration ratio may have sometimes conflicted with Italian participants' expectation, which could affect their performance in the categorization task. Finally, Italian as well as American listeners were disadvantaged by exposure to distinctly non-native sounding stimuli relative to Russian participants, who responded to stimuli with Russian-like phonetics and phonotactics.

Limitations

Among the limitations of the current study, the nature of experimental stimuli deserves some discussion. Stimuli were naturally recorded tokens, where no acoustic properties, apart from the target consonants, was artificially manipulated. As a result, all of the naturally occurring acoustic variability was preserved in the stimuli, including the stress- and context-related variability in the vowel duration, intensity, and fundamental frequency, as well as the prosodically conditioned variability in the acoustics of consonants (e.g. word-initial strengthening).

These acoustic tendencies are not uncommon across languages. For example, quantitative and qualitative reduction in unstressed vowels as compared to stressed ones is a widespread phenomenon. The dependency between vowel duration and voicing of the following consonants has also been reported for a variety of languages (Chen, 1970; Mack, 1982; Esposito, 2002; Buder & Stoer-Gammon, 2002). Articulatory and acoustic enhancement at the beginning of prosodic domains, such as the word-initial position, as opposed to reduction at the end of the domain (e.g. word-final) is also common (see, for example, Smith, 1997). Although this systematic co-variation in the acoustic properties of the stimuli could have an effect on the perception of consonant duration; the effect is expected to be consistent across the languages with similar properties. The acoustically specific nature of the stimuli certainly limits the crosslinguistic generalizability of the results and their

potential to explain crosslinguistic typology. However, given the wide range of languages with similar properties, the generalizability of the results is expected to be correspondingly high.

Another limitation was introduced by the use of stimuli which were phonetically and phonotactically more familiar to one group of participants, which could have differentially affected the performance of the experiment participants. In future research, the use of the stimuli with higher cross-language transferability could help reduce inter-language variability in the results.

Future directions

A promising avenue for exploring the effects of perception on geminate typology appears to be connected to the sonority/intensity of the target and the environment. The experimental results suggest that perception of consonant duration depends on the sonority level of the neighboring sounds: As neighbors' sonority decreases from vowels to sonorant consonants, so does the sharpness of perceptual categorization. It is also evident that sonority of targets affects the perception of duration although the direction of the effect varies across experimental studies.

However, it is not completely clear, given the currently available experimental results, how the sonority level of the neighboring sounds interacts with the sonority level of the target consonant in determining the perception of duration. At least two types of interactions are possible. In one, sonority of the environment is in the inverse relationship with sonority of targets in determining the ideal conditions for the perception of duration. In this scenario, the perfect combination of the target and the environment sonority is the one which maintains the greatest dissimilarity between the two. The best case involves the high-sonority environment (such as vowels) combined with the low-sonority targets (such as obstruents), or possibly the opposite combination: the high-sonority targets (such as liquids) in the low-sonority environment.

In the second type of interaction, sonority of the environment is in the positive relationship with sonority of targets. That is, the perceptibility of duration is boosted by both high-sonority targets and high-sonority neighbors. In this case, the ideal combination involves the high-sonority targets (such as liquids) in the high-sonority environment (such as vowels).

The present results are compatible with the second type of interaction, since high sonority consonants, such as nasals and liquids, were categorized better than stops in the high sonority vocalic environment. Pending confirmation, such an interaction would establish

intervocalic sonorants as perceptually optimal geminate types, contradicting the current typological generalization and highlighting the need for a more comprehensive survey of geminate inventories. Further research is necessary to determine the exact contribution of the sonority properties to the perception of duration in consonants and the effect that these sonority-driven differences may have on the phonological typology of duration contrasts.

Acknowledgments

I would like to thank Arto Anttila, Meghan Sumner, Jaye Padgett, Alexander Francis, and many others at Stanford University and Purdue University, as well as participants of several conferences where parts of this work were presented, for helpful discussions. The comments provided by Shigeto Kawahara and two anonymous reviewers contributed greatly to the improvement of the manuscript. Many thanks also go to the participants for their time and patience. All mistakes are my own.

Funding

This work was supported by the Ric Weiland Graduate Fellowship, Stanford University, School of Humanities and Sciences.

Notes

¹ It should be noted that the effects on sonority on geminate typology often do not hold in terms of language-internal frequency of occurrence. As pointed out by the anonymous reviewer, sonorant geminates are more frequent in Finnish than obstruent ones (e.g., Aoyama (2001) cites Finnish data where /ll/ is the second most common geminate consonant, after /tt/). Similarly, in Russian, /nn/ is among the most common geminates (Kolesnikov, 1995; Timberlake, 2004; Dmitrieva, forthcoming).

² A broad phonemic transcription is used here. In Russian, post-stress [a] reduces to schwa, while pre-stress [o] reduces to [ɐ]; [u] remains relatively unchanged in unstressed syllables (Padgett & Tabain, 2005 and references therein).

³ One of the major dialects in Liguria is Genovese (Zeneize). It should be noted that in this dialect of Italian, as in other northern dialects, gemination was lost (Ghini, 1995).

⁴ Interactions between categorical variables and three-way interactions were not specifically predicted by the research hypotheses and their inclusion in the model would significantly complicate interpretation of the results.

⁴ It should be noted that initial geminate in some languages are also believed to be moraic and could be maintained and phonetically enhanced for weight-bearing reasons, similarly to word-final geminates (Topintzi, 2008).

References

- Abramson, A. S. (1991). Amplitude as a cue to word-initial consonant length: Pattani Malay. *In Proceedings of the 12th ICPhS, Aix-en-Provence, August 1991*, 98–101.
- Abramson, A. S. (1999). Fundamental Frequency as a cue to word-initial consonant length: Pattani Malay. *In Proceedings of the 14th ICPhS, Berkeley, August 1999*, 591–594.
- Abramson, A. S. (2003). Acoustic cues to word-initial stop length in Pattani Malay. In *Proceedings of the XVth International Congress of Phonetic Sciences*, 387–390.
- Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. *Recent research in second language phonetics/phonology: Perception and production*, 231.
- Aoyama, K. (2001). *A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quantity distinctions*. Boston: Kluwer Academic Publishers
- Ardentov, B. (1979). *Fonologija sovremennogo Russkogo literaturnogo jazyka*. Kishenev: Shtiintsa [Phonology of contemporary literary Russian language].

Armosti, S. (2009). *The Phonetics of Plosive and Affricate Gemination in Cypriot Greek*.

PhD thesis, University of Cambridge.

Arvaniti, A. and Tserdanelis, G. (2000). On the phonetics of geminates: evidence from

Cypriot Greek. *In Proceedings of 6th International Conference on Spoken*

Language Processing, 2, 559–562.

Avanesov, R. I. (1984). *Russkoje Literaturnoje Proiznoshenije*. Prosveshchenie, Moskva.

[Russian Literary Pronunciation].

Baum, S. R. (1992). The influence of word length on syllable duration in aphasia: Acoustic

analyses. *Aphasiology*, 6(5), 501-513.

Beddor, P., & Evans-Romaine, D. (1992). Acoustic-perceptual factors in nasal place

assimilation. *JASA*, 91, 24-73.

Bell, A. J. (2003). Gemination, degemination and moraic structure in Wolof. *Working*

Papers of the Cornell Phonetics Laboratory, 15, 1–68.

Blevins, J. (2004). *Evolutionary Phonology*. Cambridge, UK: Cambridge University Press.

- Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [computer program]. URL <http://www.praat.org/>.
- Borg, A. (1997). Maltese phonology. In A. S. Kaye (Ed.), *Phonologies of Asia and Africa* (pp. 245–285). Winona Lake, IN: Eisenbrauns.
- Bradley, T. (2001). *The phonetics and phonology of rhotic duration contrast and neutralization*. PhD thesis, The Pennsylvania State University.
- Buder, E. H., & Stoel-Gammon, C. (2002). American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *The Journal of the Acoustical Society of America*, 111(4), 1854-1864.
- Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics*, 83, 97–115.
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22(3), 129-159.
- Cho, T. and Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, 37, 466–485.

- Cowell, M. (1964). *A reference grammar of Syrian Arabic. Based on the dialect of Damascus*. Washington, D.C.: Georgetown University Press.
- Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods, 15*(3), 339–362.
- Davis, S. (1994). Geminate consonants in moraic phonology. *Proceedings of the West Coast Conference on Formal Linguistics, 13*, 32–45.
- Davis, S. (2011). Geminates. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice, (Eds.), *The Blackwell Companion to Phonology*, volume 2 (pp. 837–859). Malden, MA and Oxford, UK: Wiley-Blackwell.
- Delattre, P.C., Liberman, A.M., & Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. *JASA, 27*, 769-73.
- Dmitrieva, O. (forthcoming). Production of geminate consonants in Russian: Implications for typology. In H. Kubozono (ed.) *The Phonetics and Phonology of Geminate Consonants*. Oxford: Oxford University Press.

Erwin, W. M. (1963). *A short reference grammar of Iraqi Arabic*. Washington, D.C.: Georgetown University Press.

Esposito, A. (2002). On vowel height and consonantal voicing effects: Data from Italian. *Phonetica*, 59(4), 197-231.

Esposito, A., & Di Benedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustical Society of America*, 106(4), 2051-2062.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 232–276). Cambridge, UK: Cambridge University Press.

Fujimura, O., Macchi, M., & Streeter, L. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech*, 21, 337–346.

Ghini, M. (1995). The coronal-velar relationship in Genovese Ligurian. *Toronto Working Papers in Linguistics*, 14, 55–72.

Goedemans, R. 1998. *Weightless segments*. The Hague: Holland Academic Graphics.

Gordon, M. (2002). A phonetically driven account of syllable weight. *Language*, 78(1), 51-80.

Gordon, M. (2006). *Syllable Weight: Phonetics, Phonology, Typology*. New York: Routledge.

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, 54(3), 383-394.

Hansen, B. (2012). *The perceptibility of duration in the phonetics and phonology of contrastive consonant length*. PhD thesis, The University of Texas at Austin.

Harrison, S. P. (1995). Marshallese. In D. T. Tryon (Ed.), *Comparative Austronesian Dictionary*, Part 1, Fascicle 2 (pp. 895–903). Berlin, New York: Mouton de Gruyter.

Hawkins, J. A., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 280-317). Blackwell.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20, 53–306.

Hayes, Bruce. 1980. *A metrical theory of stress rules*. Ph.D. thesis, MIT. Published 1985. New York: Garland.

Hume, E., & Johnson, K. (2001). A Model of the interplay of speech perception and phonology. In E. Hume, & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 251–277). Academic Press.

Hura, S. L., Lindblom, B., & Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech*, 35, 59–72.

Hutchins, S., Gosselin, N., & Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in psychology*, 1, 236.

IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Jaeger, J. (1978). Speech aerodynamics and phonological universals. *Proceedings of BLS*, 4, 311–329.

- Kasatkin, L. L., & Choj, M. C. (1999). *Dolgota/kratkost' Soglasnogo na Meste Sochetanij Dvuh Bukv v Sovremennom Russkom Literatunom Jazike*. Moskva: Dialog-MGU. [Duration of the doubly-spelled consonants in the Contemporary Standard Russian].
- Kato, H., Tsuzaki, M., & Sagisaka, Y. (1997). Acceptability for temporal modifications of consecutive segments in isolated words. *JASA*, *101*, 2311–2322.
- Kawahara, S. (2007). Sonorancy and geminacy. *University of Massachusetts Occasional Papers in Linguistics*, *32*, 145–186.
- Kawahara, S., & Pangilinan, M. (in press). Spectral continuity, amplitude changes, and perception of length contrasts. In H. Kubozono (Ed.) *Phonetics and Phonology of Geminate Consonants*. Oxford: Oxford University Press. Retrieved on 08/27/2016 from http://user.keio.ac.jp/~kawahara/pdf/GC_KawaharaPangilinan_Revision.pdf
- Kotzor, S., Molineaux, B. J., Banks, E., & Lahiri, A. (2016). “Fake” gemination in suffixed words and compounds in English and German. *The Journal of the Acoustical Society of America*, *140*(1), 356–367.
- Kolesnikov, N. P. (1995). *Slovar' Slov s Dvojnymi Soglasnymi*. Izdatel'stvo Rostovskogo Universiteta, Rostov-na-Donu. [Dictionary of words with double consonants].

- Kraehenmann, A. (2011). Initial Gemimates. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1124–1146). Malden, MA & Oxford: Wiley-Blackwell.
- Kraehenmann, A. (2001). Swiss German stops: Gemimates all over the word. *Phonology*, 18, 109–145.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America*, 51(6B), 2018-2024.
- Lieberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48(4), 839-862.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala, & J. Jaeger (Eds.), *Experimental Phonology* (pp. 13-44). Orlando, FL: Academic Press.
- Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America*, 71(1), 173-178.

- Mahootian, S. (1997). *Persian*. London, UK: Routledge.
- Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus intensity and the perception of duration. *Journal of Experimental Psychology: Human perception and performance*, 37(1), 303.
- Matusevich, Margarita Ivanovna. (1976). *Sovremennij Russkij Jazyk, Fonetika. Prosveshchenie, Moskva. [Contemporary Russian, Phonetics]*.
- McCarthy, K. M., Mahon, M., Rosen, S., & Evans, B. G. (2014). Speech Perception and Production by Sequential Bilingual Children: A Longitudinal Study of Voice Onset Time Acquisition. *Child development*, 85(5), 1965-1980.
- Muller, J. (2001). *The phonology and phonetics of word-initial geminates*. PhD thesis, Ohio State University.
- Ohala, J. (1990). The phonetics and phonology aspect of assimilation. In J. Kingston, & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 258–275). Cambridge, UK: Cambridge University Press.
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54, 1235–1247.

Padgett, J. (2003). Systemic contrast and Catalan rhotics. Ms., University of California, Santa Cruz.

Padgett, J., & Tabain, M. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica*, 62(1), 14-54.

Pajak, B. (2010). Contextual constraints on geminates: The case of Polish. *Proceedings of BLS*, 35, 269-280.

Pajak, B. (2013). Non-intervocalic geminates: Typology, acoustics, perceptibility. In L. Carroll, B. Keffala, & D. Michel (Eds.), *San Diego Linguistics Papers 4* (pp. 2-27). San Diego, CA: UC San Diego.

Panov, M. V. (1979). *Sovremennij Russkij Jazyk, Fonetika. Vischaja Shkola, Moskva.* [The Contemporary Russian Language, Phonetics].

Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica*, 56(3-4), 135-157.

Pickett, J. M., and Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech* 3, 11-17.

- Plourde, M., Meilleur-Wells, G., Gamache, P. L., Dionne, G., & Grondin, S. (2008). Relative sensitivity to filled versus empty intervals depends on methods: Evidence from developmental data. *Proceedings of Fechner Day*, 24(1), 93-98.
- Podesva, R. (2000). Constraints on geminates in Buginese and Selayarese. *Proceedings of the West Coast Conference on Formal Linguistics, volume 19*, 343–356.
- Podesva, R. (2002). Segmental constraints on geminates and their implication for typology. Talk presented at the 76th Annual Meeting of the Linguistic Society of America, San Francisco.
- Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from <http://www.pstnet.com>.
- Pycha, A. (2010). A test case for the phonetics-phonology interface: Gemination restrictions in Hungarian. *Phonology*, 27, 119–152.
- Rammsayer, T. H. (2010). Differences in duration discrimination of filled and empty auditory intervals as a function of base duration. *Attention, Perception, & Psychophysics*, 72(6), 1591-1600.

- Ridouane, R. (2007). Gemination in Tashlhiyt Berber: An acoustic and articulatory study. *Journal of the International Phonetic Association*, 37, 119–142.
- Ringen, C., & Vago, R. M. (2011). Geminates: Heavy or long? In C. Cairns, & E. Raimy (Eds.), *Handbook of the Syllable* (pp.155–169). Leiden: Brill.
- Rogers, D. & D’Arcangeli, L. (2004). Italian. *Journal of International Phonetic Association* 34, 117–121.
- Santi, A., Miki, A., Hornyak, S., & Eidse, J. (2005). The perception of empty and filled time intervals by rats. *Behavioural processes*, 70(3), 247-263.
- Schwartz, J-L., Boe, L-J., Vallee, N., & Abry, C. (1997b). Major trends in vowel system inventories. *Journal of Phonetics*, 25, 233–253
- Schwartz, J-L., Boe, L-J., Vallee, N., & Abry, C. (1997a). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, 25, 255–286.
- Seiler, G. (2009). Sound change or analogy? Monosyllabic lengthening in German and some of its consequences. *Journal of Comparative German Linguistics*, 12, 229–272.

- Smith, C. L. (1997). The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics*, 25(4), 471-500.
- Spaelti, P. (1994) Weak Edges and Final Geminates in Swiss German, in M. Gonzalez (Ed.), *Proceedings of NELS 24* (pp. 573-588). Amherst, MA: GLSA, University of Massachusetts.
- Steriade, D. (1997). Phonetics in phonology: The case of laryngeal neutralization. Ms., UCLA.
- Steriade, D. (2001). Directional asymmetries in place assimilation. In E. Hume, & K. Johnson (Eds.), *The Role of Speech Perception in Phonology* (pp. 219–250). New York, NY: Academic Press.
- Taylor, M. (1985). Some patterns of geminate consonants. *University of Chicago Working Papers in Linguistics*, 1, 120–129.
- Thurgood, G. (1993). Geminates: A cross-linguistic examination. In J. A. Nevis, G. McMenamin, & G. Thurgood, G (Eds.), *Papers in Honor of Frederick H. Bregelman on the Occasion of the Twenty-fifth Anniversary of the Department of Linguistics, (CSU) Fresno* (pp. 129–139). Fresno, CA: Department of Linguistics, California State University Fresno.

- Timberlake, A. (2004). *A Reference Grammar of Russian*. Cambridge: Cambridge University Press.
- Topintzi, N. (2008). On the existence of moraic onsets. *Natural Language and Linguistic Theory*, 26, 147–184.
- Turk, A. (1992). The American English flapping rule and the effect of stress on stop consonant duration. *Working Papers of the Cornell Phonetics Laboratory*, 7, 103–133.
- Turk, A. and White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27(2), 171–206.
- Turk, A. E. and Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, 28, 397–440.
- Wearden, J. H., Norton, R., Martin, S., & Montford-Bebb, O. (2007). Internal clock processes and the filled-duration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3), 716.

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

Appendix A: Results of the follow-up models

The Stress/Manner model

Table 3: Significant effects and interactions in the Stress/Manner model

<i>Factor</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Manner	36.247	4	23,849	.000
Stress	8.019	2	23,849	.000
Language	10.999	2	23,849	.000
Duration	202.767	1	23,849	.000
Duration*Manner	16.777	4	23,849	.000
Duration*Stress	3.572	2	23,849	.028
Language*Language	4.802	2	23,849	.008

Table 4: Terms with significant coefficients in the Stress/Manner model

	<i>Model Term</i>	<i>Coeff-ent</i>	<i>Std. Error</i>	<i>p</i>	<i>Exponented Coeff-ent</i>
	Intercept	0.676	0.261	.010	1.966
Reference: /s/	/l/	1.789	0.253	.000	5.985
	/n/	1.782	0.285	.000	5.944
	/t/	-1.572	0.215	.000	0.208
Reference: elsewhere stress	Preceding stress	0.543	0.142	.000	1.721

Reference: <i>English</i>	Russian	1.739	0.497	.000	5.692
	Italian	1.611	0.404	.000	5.009
Duration	Duration	0.029	0.003	.000	1.030
Reference: <i>Duration*/s/</i>	Duration*/d/	-0.009	0.002	.000	0.991
	Duration*/t/	-0.016	0.002	.000	0.984
Reference: <i>Duration*elsewhere stress</i>	Duration*Preceding stress	0.002	0.001	.008	1.002
Reference: <i>Duration*English</i>	Duration*Russian	0.013	0.004	.003	1.013

The Environment model

Table 5: Significant effects and interactions in the Environment model

<i>Factor</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Environment	4.124	1	6,407	.000
Language	2.782	2	6,407	.062
Duration	245.740	1	6,407	.000
Duration*Environment	4.286	1	6,407	.038
Language*Language	3.506	2	6,407	.030

Table 6: Terms with significant coefficients in the Environment model

	<i>Model Term</i>	<i>Coeff-ent</i>	<i>Std. Error</i>	<i>p</i>	<i>Exponented Coeff-ent</i>
	Intercept	-0.174	0.242	.472	0.840
Reference: <i>preconsonantal</i>	Intervocalic	0.274	0.135	.042	1.315
Reference: <i>English</i>	Russian	0.834	0.376	.026	2.303
Duration		0.026	0.002	.000	1.026
Reference: <i>Duration*preconsonantal</i>	Duration*Intervocalic	0.004	0.002	.038	1.004

Reference: <i>Duration*English</i>	Duration*Russian	0.016	0.006	.008	1.016
---------------------------------------	------------------	-------	-------	------	-------

The Position model

Table 7: Significant effects and interactions in the Position model

<i>Factor</i>	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Position	11.249	1	6,031	.001
Duration	356.869	1	6,031	.000
Duration*Position	3.212	1	6,031	.073
Language*Language	2.641	2	6,031	.071

Table 8: Terms with significant coefficients in the Position model

	<i>Model Term</i>	<i>Coeff-ent</i>	<i>Std. Error</i>	<i>p</i>	<i>Exponented Coeff-ent</i>
	Intercept	-0.723	0.286	.011	0.485
Reference: <i>final</i>	Initial	0.591	0.176	.001	1.805
Duration		0.026	0.002	.000	1.026
Reference: <i>Duration*preconsonantal</i>	Duration*Intervocalic	0.003	0.002	.073	1.003