Laboratory Phonology
Journal of the Association for Laboratory Phonology

Mitterer, H. 2018 The singleton-geminate distinction can be rate dependent: Evidence from Maltese. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9(1): 6, pp. 1–16, DOI: https://doi.org/10.5334/labphon.66

JOURNAL ARTICLE

The singleton-geminate distinction can be rate dependent: Evidence from Maltese

Holger Mitterer
Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida, MT
holger.mitterer@um.edu.mt

Many languages distinguish short and long consonants, or singletons and geminates. The primary acoustic correlate of this distinction is the duration of the consonants. Given that the absolute duration of speech sounds varies with speech rate, the question rises to what extent the category boundary between singletons and geminates is sensitive to the overall speech rate (i.e., rate normalization). Next to rate normalization, there are two other possible explanations how singletons and geminates might be distinguished. First, it has been suggested that despite variation in absolute duration, the two categories remain distinct; that is, even in fast speech, geminates seldom take on durations that would be typical of singletons at slow speech rates. Second, it has been suggested that, with higher speech rate, both the duration of consonants and vowels shrink, so that the duration ratio of consonant and adjacent vowel is a rate independent cue for the singleton-geminate distinction. Using production and perception data from Maltese, we show that, first, the singleton-geminate distinction is endangered by speech-rate variation and, second, consequently undergoes speech-rate normalization.

Keywords: geminates; speech-rate normalization; Maltese

1. Introduction

Geminate consonants are consonants which are longer than their regular, singleton counterparts and this distinction is phonemic in many languages (e.g., in Maltese, *kiser* = 'he broke,' *kisser* = 'he smashed'). It is well established that duration is the most potent cue overall to distinguish singleton and geminates (Hankamer & Lahiri, 1988; Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009; Yoshida, de Jong, Kruschke, & Päiviö, 2015). For speech sounds that are distinguished by duration, the question rises how the distinction can be maintained despite variation in speech rate. With variation in speech rate, the absolute duration of speech sounds gets longer or shorter. How can listeners make a distinction between 'short' and 'long' consonants in such circumstances? It has been argued that the perception of duration in speech is often rate-dependent (Newman & Sawusch, 1996, 2009; Port, 1979; Reinisch & Sjerps, 2013; Summerfield, 1981), so that a 'medium' duration is interpreted as 'contrastively' long if the surrounding rate is fast but as short if the surrounding rate is slow (Bosker, 2017). In this paper, the question is asked whether this sort of rate dependency also influences the singleton-geminate distinction, using Maltese, a Semitic language that makes use of consonant quantity (for an overview, see Galea, 2016).

At first sight, it might seem foolish to ask this question; why would it not? However, there are two reasons why the singleton-geminate distinction might not be rate-dependent. First, Nakai and Scobbie (2016) argued that rate-normalization might not be necessary for some contrasts because the category boundary does not necessarily shift with rate. For the case of English stops, they showed that what changes with rate is mostly the amount of aspiration in voiceless stops, while the VOT in voiced stops is hardly affected. Voiced stops, even at a slow rate of speech, hence seldom have a VOT that makes them similar to aspirated, unvoiced, stops. A similar finding is reported for Icelandic stops by Pind (1995). Such findings are not restricted to VOT in stops, as Miller and Baer (1983) found that, for the distinction of /b/ and /w/ in English, the transition duration changes with rate mainly for /w/ but not for /b/ (see also Port, 1979, for a distinction between flap and stop). This suggests that generally, when duration is critical for a phonemic distinction, rate strongly affects only one member of the distinction and, consequently, the category boundary is not necessarily dependent on speech rate. Critically for the issue at hand, Arvanti (1999) found that there is indeed little overlap between singleton and geminate categories in Cypriot Greek despite variation in speaking rate. Consequently, there would be no need for rate normalization. Note, however, that others have reported overlap in absolute duration for quantity distinctions depending on speaking rate (see, e.g., Hirata, 2004, for Japanese vowels).

Another reason why rate normalization might not be necessary for the singleton-geminate distinction stems from another approach that argues that there is a distinction between rate-normalization 'proper' and the use of multiple cues for a given distinction. Even when there is overlap between the duration of singleton and geminate consonants caused by rate, it might be sufficient to take into account the duration of the neighboring vowels rather than the ambient speech rate (Idemaru & Guion-Anderson, 2010; E. R. Pickett, Blumstein, & Burton, 1999; Pind, 1995). That is, the duration ratio between adjacent vowel and consonant is sufficient to correctly categorize an utterance as singleton (VCV) or geminate (VC:V) independent of rate, because as the consonant gets longer, so does the vowel, leading to a rate-independent ratio. This approach dovetails well with recent approaches in psychology that view rate-normalization not as rate-normalization per se, but rather as a by-product of using multiple cues for a given distinction (McMurray & Jongman, 2011). Based on this approach, Toscano and McMurray (2012) tested the time course of cue utilization in speech perception using eye-tracking (cf. McMurray, Clayards, Tanenhaus, & Aslin, 2008; Mitterer & Reinisch, 2013). They tested the rate dependency of VOT and found that the duration of the adjacent vowel is used as an independent cue rather than as a moderating effect on the VOT cue, as would be predicted by a rate-normalization account. In line with the assumption that speech rate per se may not influence segment perception, Newman and Sawusch (1996) found that *distal* rate does not influence segment decisions. In these experiments, the speech rate in a carrier sentence was manipulated up to 400 ms, but was left unchanged around the critical segment.

Regarding the singleton-geminate distinctions, Pind (1986) found strong effects of the neighboring vowel, while external speaking rate had only a minor influence on categorization of a given stimulus; boundaries shifted only by 3 ms, which was estimated to be below the JND with a base duration of about 100 ms. On the other hand, Hirata and Lambacher (2004) found that excising words with long and short vowels from their context, or putting them into a carrier phrase with a mismatching speaking rate, led to large number of misidentifications. In this case, however, the external context was still rather close to the critical vowel, being separated only by one consonant. In the speech-perception literature, an effect of rate is only considered distal when there is at least a full syllable between the target and a segment with a manipulated speech rate (Heffner, Newman, & Idsardi, 2017). The available evidence would hence suggest that external speaking rate has only a small influence on the perception of the singleton-geminate distinction.

While there is a scarcity of evidence that distal rate may influence segment perception, there is clear evidence that distal rate can influence speech segmentation (Dilley & McAuley, 2008; Dilley & Pitt, 2010; J. M. Pickett & Decker, 1960; Reinisch, Jesse, &

McQueen, 2011). Consider a phrase as *Canadianoats* [kəneɪdiənəʊts], which could mean "Canadian oats" or "Canadian notes." For such cases, listeners take rate into account and more often perceive the phrase "Canadian notes" in fast speech; that is, they accept a shorter duration as being long (Reinisch et al., 2011), in line with a contrastive duration perception (Bosker, 2017). This state of the literature may give rise to the impression that distal speech rate only matters for speech segmentation but not for speech segments, an issue that will resurface in the General Discussion.

With this potential generalization about rate effects on segments and segmentation, there is ample reason to doubt that there is rate-dependent perception of the singleton-geminate contrast. There are nevertheless also recent indications that perception can be rate dependent. Reinsich and Sjerps (2013) tested the perception of the Dutch vowel contrast /ɑ /-/a/, which differs both in spectral and temporal properties, and found with eye-tracking that preceding rate and spectral information triggered immediate context effects in segment perception, which in turn argues for rate (and spectral) normalization. However, the question whether there is rate-normalization or not may be ill-posed at a general level. Recent evidence suggests that rate normalization may be important for some contrast but not for others (Heffner et al., 2017). It is therefore worthwhile to test whether a quantity distinction might be rate-dependent.

The question whether the perception of a singleton-geminate distinction is rate-dependent raises two issues. First, is the category boundary endangered by rate variation in production? That is, do geminates produced at a fast rate ever get so short that their absolute duration is in the same range as the duration of singletons produced at a slow rate? To answer this, we analyzed a corpus of utterances containing singleton-geminate minimal pairs produced in a sentence context. We used the same methods as used by Nakai and Scobbie (2016) for VOT duration in English stops, by testing whether categorization accuracy improves when speech rate is taken into account. Second, do listeners consider (distal) speech rate when categorizing sounds as singleton versus geminate? This was investigated by means of perception experiments in which the distal speech rate was manipulated.

2. Production study

2.1. Method

To test whether the perception of the singleton-geminate boundary might be usefully rate-dependent, we made use of data from a production study in which Maltese participants produced Maltese verbs in their first and second *binyam* form (Mitterer, 2018). Maltese, like other Semitic languages, has a rich verb morphology, both inflectional and derivational. Semitic[1] verbs are based on tri-consonantal roots (using the standard example, *k-t-b*, for writing) which can be used to generate verbs in the first form (*kiteb*, 'he wrote') and in the second form, which usually has a causative or intensive meaning (*kitteb*, 'he wrote regularly'). Just as in these examples, the first and second form of a verb (in 3rd male singular + past) forms a minimal or near-minimal pair that is distinguished by the quantity (singleton vs. geminate) of the middle consonants. This allows us to generate many minimal pairs.[2] Here, utterances from a corpus were used, in which these verbs were elicited by picture primes and participants had to guess the sentence. For instance, the participants would see a cartoon character (whose name has been established as, e.g., 'Daniel' in a training phase) a root (e.g., *r-q-d*, 'sleep') and another object (e.g., a sofa) which should be put together in a sentence (e.g., "Daniel slept on the sofa"). Each

[1] Not all Maltese verbs are Semitic in origin; there are verbs that stem from Italian (e.g., *ikkanta*, 'to sing') and English (e.g., *iċċekja*, 'to check') which do not partake in the derivational verb morphology outlined here.

[2] Maltese allows all consonants to geminate, and also allows geminates in word-initial and -final position (Galea, 2016).

participant saw these forms in sets of five (including fillers in which the plural present tense was required), first having to guess the sentence, and then, with a repetition of these five items remembering the sentence.

There were 36 minimal or near-minimal pairs, all the form CVC(C)VC, that were elicited from fourteen speakers.[3] Three of these forms form only near-minimal pairs, because the filler vowels added to the root consonants differ between first and second form (e.g., *weħel-waħħal.* 'he got stuck' and 'he attached'). Given the non-reading nature of the production task, not all prompts led to the production of the target form. From the 2016 total trials, 1355 could be used for this analysis. On average, there were 47 singletons and 50 geminate items per speaker. If further subdivided by segments, there were 10 to 14 items per speaker in each cell defined by segment and quantity.

Duration of the consonants was estimated using forced alignment. Given that there was some variation in the exact wording of the sentence (e.g., participants misremembering the name of a cartoon character, or interpreting the picture of a little girl as 'his little sister'), the analysis of duration focused on the target form. The start and end of the target form was marked by hand and then forced alignment was achieved using Praatalign (Lubbers & Torreira, 2013), which makes use of the phones of the Munich AUtomatic Segmentation System, using the language-independent mode in which best trained phones from all languages with training data are used (Strunk, Schiel, & Seifart, 2014).

There was no explicit manipulation of rate in this production task. This in contrast with Arvaniti (1999), who asked speakers to speak at a normal pace or faster. In the current case, it became apparent during the coding process that the speakers were internally consistent but different from each other in how fast they produced these sentences naturally. That is, one way to estimate the role of rate for the category boundary is to estimate it for the speakers separately and then for the whole sample. If the differences in rate between speakers mattered, the accuracy of categorization based on measured consonant duration should be significantly higher with different boundaries for each speaker. Moreover, the boundaries should be at shorter durations for speakers with an above average speaking rate.

To test this, the average speaking rate per speaker was estimated as follows. For all useable utterances, the speaking rate was estimated using the method as proposed and implemented by de Jong and Wempe (2009). These estimates were hand-corrected for, first, the number of syllables uttered and potential misclassifications of stop closures as pauses (which frequently occurred for geminate stops). This provides an estimated average syllable duration for each utterance. However, this average syllable duration is still influenced by the target word itself. To achieve an estimate of speaking rate that was not confounded with any influences of the target word and other extraneous influences, such as the number of pre-pausal lengthenings, average syllable duration per utterance was predicted by a linear-mixed effect model with two fixed effects, the number of pauses and whether the sentence prompt was seen for the first or second time, and two random intercepts, speaker and item. Adding random slopes for the fixed effects did not increase fit even with a relatively anti-conservative criterion ($p > 0.2$). The model unsurprisingly revealed effects of number of pauses, with longer duration if there are more pauses (b = 0.005, t(1329) = 6.362, $p < 0.001$) and shorter average syllable durations when a

[3] A full description of the corpus generation and all sentence materials can be found in Mitterer (2018). The main objective of the corpus generation was to investigate the role of secondary cues to gemination for oral versus glottal segments, a research question orthogonal to the current one.

prompt is seen for a second time (b = –0.006, t(1042) = –6.109, $p < 0.001$). Critically, the model provides a random effect for speaker, which is an estimate whether a speaker has an above or below average speaking rate, while other extraneous influences, such as whether the item was a singleton or geminate, are controlled for.

Optimal boundaries were estimated following the procedure of Nakai and Scobbie (2016), who followed the procedure proposed by Miller, Green, and Reeves (1986). The same algorithm was used here to estimate the optimal boundary based on measured consonant duration. The algorithm estimated for the whole range of observed durations (in their case, VOT duration, in the current case, segment duration) how many items are correctly categorized based on their duration if the boundary is set to any duration value in that range. That is, if the longest segment is 200 ms long, and the shortest is 60 ms long, it is tested for each duration in this interval (using a step size of 1 ms) how many tokens would be correctly classified if the boundary is assumed at this duration. The optimal boundary is estimated as the duration at which the likelihood of correct categorization is maximal.

2.2. Results

Figure 1 shows the results of finding the optimal boundary for each of the 14 speakers and the complete sample in separate panels. The classification accuracy was 81.8% for the boundary (estimated at 119 ms) from the complete sample but 85.8% if estimated for each speaker separately (estimated boundaries ranging from 80 to 130 ms), which is a significant improvement (using a Chi-square test with Yates continuity correction, $X^2(1) = 7.650$, $p = .005$). Maybe more importantly, there was a strong correlation between each speaker's boundary and his or her estimated speaking rate estimated from the linear mixed-effect model described above (r = 0.74, $p < 0.005$). That is, speakers who had an above average syllable duration (i.e., a slow speaking rate) also had a category boundary between singleton and geminates that was above average.

Note, however, that the overall classification accuracy is rather low with maximally 85% correct. We therefore also considered segment identity and found an optimal category boundary for each combination of speaker and segment and compared this with an optimal category boundary for each segment. This led to higher classification accuracies of 91.4% correct classifications with speaker considered and 85.6% without taking speaker into account (which still is a significant difference, $X^2(1) = 20.12$, $p < 0.001$). Undoubtedly, higher accuracies could be reached if even carrier word is taken into account; however, the current data then become too sparse to allow a meaningful estimation of an optimal category boundary.

Since previous research found that often only one member of a category distinction is affected by rate, we correlated (over speakers) the estimated speaking rate with both the mean geminate and singleton durations and obtained similar correlations (singleton: r = 0.625, $p < 0.05$; geminate: r = 0.761, $p < 0.01$, Fisher z-test for a difference between these correlations, z = 0.62, $p = 0.54$). This indicates that the duration of both singletons and geminates in Maltese varies with speaking rate.

It is also worth considering what the classification accuracy would be when the consonant/vowel duration ratio is used instead of the consonant duration to estimate the optimal boundary, given that the C/V ratio has been proposed as an higher-order invariant (Pind, 1995). Therefore, the optimal boundary was also determined using ratio of the consonant duration and the preceding vowel, the following vowel, and the average duration of these two vowels. This gave rise to a classification accuracy of 69.7%, 74.5%,

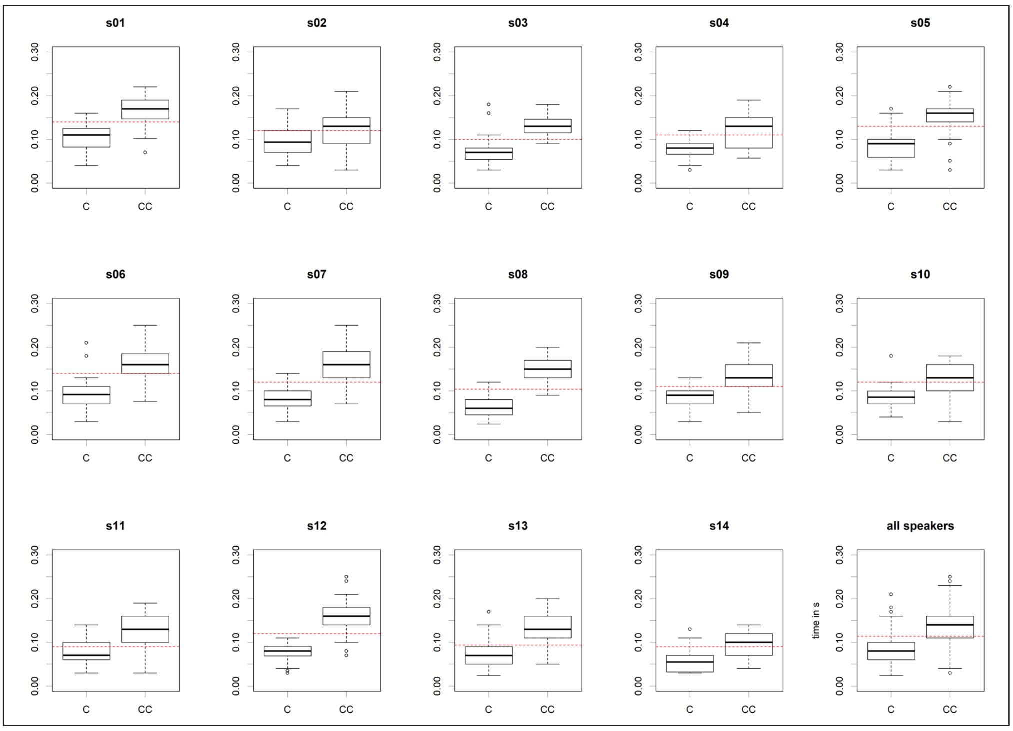

Figure 1: Boxplot of durations of singletons (C) and geminates (CC) for each speaker and the complete sample (lower right panel). The black lines indicate the median. The boxes' lower and upper border are based on the 25th and 75th percentile and hence show where 'middle' 50% of the data are to be found. The whiskers show minimum and maximum values that are within 1.5 times the inter-quartile range of the 25th and 75th percentile and any values exceeding these values are represented by dots. The red dashed lines show the estimated optimal category for each boxplot based on the same algorithm used by Nakai and Scobbie (2016).

and 75.9%, respectively. This is worse than the 81.8% based on duration alone, and hence also significantly less than the classification accuracy with habitual speech rate considered.

2.3. Discussion

The results indicate that in contrast to VOT in English (Nakai & Scobbie, 2016) and consonant duration for quantity in Cypriot Greek (Arvaniti, 1999), the optimal boundary between singleton and geminates in Maltese varies with speaking rate and does so considerably, given the large range (80–130 ms) of estimated boundaries. Moreover, the data also indicate that both categories are similarly affected by speech rate. If only geminates would be influenced by rate, there would be no danger for the category boundary, since the singletons would remain relatively short even at slow speaking rates. This gives rise to the possibility that the singleton-geminate distinction in perception is sensitive to speech rate.

The data also are not in line with the assumption that the consonant-vowel duration ratio may be a higher-order invariant cue for Maltese quantity. It is, however, important to note that the opposite finding would not have been diagnostic with regard to the two different accounts for contrast maintenance, speech-rate normalization versus higher-order invariants. Speech rate normalization argues that the duration of a consonant is evaluated against the average duration of the surrounding segments, while the assumption of the higher-order invariant argues that only the duration of the two adjacent vowels matter. Because these measures will often be correlated, the finding that the vowel duration suffices is not showing that speech rate is not being taken into account in perception.

As this shows, simply measuring the cues in a speech corpus by itself is principally not a good way to measure the perceptually relevant features of a contrast, due to the correlational nature of any set of measurements. The best evidence is provided by experimental data based on manipulating cues independently.

Therefore, a perception experiment was conducted in which participants performed a two-alternative forced choice (2AFC) task, deciding whether the critical word in a sentence contained either a singleton or a geminate (e.g., is the critical word *wasal* or *wassal*, 'to arrive' and 'to bring,' respectively). These words were presented in a sentence context (*Anna tipprova ma tuzax il kelma …. f'dan il-kaz*, 'Anna tried to not use the word … in this case'), and only the sentence context was varied in rate, but not the critical word itself. We used a carrier phrase in which the critical word was not utterance-final so that listeners would not have to account for any utterance-final lengthening. The two accounts for how the singleton-geminate contrast is adjusted for variation in speech rate predict different outcomes. The account based on a higher-order invariant predicts that there should be no effect—or maximally a small effect (cf. Pind, 1986)—of the ambient rate, since the (then primary) consonant-vowel ratio is constant. The speech-rate normalization account, in contrast, predicts that the ambient rate should influence the perception of the critical word.

3. Experiment 1

3.1. Method

3.1.1. Participants

Sixteen native speakers of Maltese participated in the experiment for pay. They were aged between 19 and 41 years and 11 of them were female. They all reported to have learned Maltese before English[4] and used Maltese for at least 50% of their daily interactions both during childhood and adolescence. They all filled in an informed-consent form before the experiment started.

3.1.2. Materials

A female native speaker of Maltese produced multiple renditions of the carrier sentence *Anna tipprova ma tuzax il kelma …. f'dan il-kaz,* 'Anna tried to not use the word … in this case,' in which the empty space was filled by a member of three singleton-geminate minimal pairs: *qata'-qatta'*, 'he cut – he chopped up,'[5] *rikeb-rikkeb*, 'he rode – he gave a ride,' *wasal-wassal* 'he arrived – he brought.' From these utterances, the typical duration of the singleton and geminates in these sentences was estimated and rounded to the nearest value for which the modulus by ten was zero (*qata'-qatta'*: 80 vs 180 ms, *rikeb-rikkeb*: 100 vs 180 ms, *wasal-wassal*: 100 vs 200 ms). Duration continua were then generated by extracting a geminate utterance that was slightly longer than the typical duration for a geminate and then cutting back the duration of the consonant in five steps for each of the three minimal pairs, starting from the typical geminate duration up to the typical singleton duration. This gives rise to 15 stimuli (five durations for three continua).

To generate target sentences, a different sentence was used for each target continuum to prevent coarticulatory mismatches especially between the final word of the target-preceding part (*kelma*) and the following target word. From three selected sentences, the target-preceding and -following parts were extracted. All sounds, that is, precursors (*Anna tipprova ma tuzax il kelma*), target minimal-pair continua, and following contexts (*f'dan il-kaz*) were then rate manipulated using the PSOLA algorithm in Praat (Boersma,

[4] Malta is officially bilingual (English/Maltese) but Maltese is the primary language spoken in social situations, though English is the official language at University. Only a minority of speakers (<10%) is more proficient in English than in Maltese.

[5] The first and second form of a verb is, with a few exceptions, only a minimal pair distinguished by consonant quantity for the 3rd male past tense form, since this so-called 'mama' form of the verb does not contain any affixes. The affixes are added in a slightly different way for the first and second form (e.g., *rkibt*, 'I rode' but *rikkibt*, 'I gave a ride'). Note also that Maltese verbs do not have an infinitive form.

2001). The carrier phrases (i.e., precursors and following contexts) were either sped up or slowed down by 20%. Targets were decelerated by 10% and then again accelerated by 11.1% (i.e., 1/0.9) to retain their original duration. This ensured that both the carrier sentence and the targets were speech signals derived from a PSOLA resynthesis. This was done because PSOLA can introduce slight artefacts which might render untreated targets to stand out from PSOLA-manipulated carrier phrases. In these target syllables, the first syllable had a duration of 170 ms for *was(s)sal*, *96 ms* for *rik(k)eb*, and 153 ms for *qat(t)a'*. After the rate manipulation, the stimuli were corrected slightly at the splicing points (<5 ms) so that phases of the glottal cycles appeared continuous in the recombined stimuli (i.e., the precursors all ended on a major positive going zero-crossing and the target, if voiced in the onset, started after the major positive going zero crossing). After this correction, each member of the three target continua was concatenated with the slow and fast versions of the carrier phrase, giving rise to 30 stimuli (2 rates × 3 continua × 5 durations per continua).

3.1.3. Procedure

After reading the informed-consent form, participants were placed in front of a 19-inch monitor driven by a standard PC computer and placed in a sound-attenuated booth at the Cognitive-Science Lab at the University of Malta. Experimental sessions were controlled using PsychoPy (Peirce, 2007). An on-screen instruction explained the 2AFC procedure to the participants. Answer options were presented as the full written words presented on the right or left lower half of the screen (e.g., *rikeb* and *rikkeb*, with the singleton option always presented on the left). The different continua were presented intermixed, that is, on a given trial the minimal pair might be *rikeb-rikkeb*, and on the next trial it might be *qata'-qatta'*, *wasal-wassal*, or a repetition of *rikeb-rikkeb*. Each of the 30 stimuli was presented 10 times to each participant, randomized in such a fashion that participants listened to 10 permutations of the 30 stimuli. That is, the whole range of stimuli were presented once before the first stimulus was presented for the second time.

Participants responded by pressing the left or right arrow button on a keyboard. After their reaction, their choice was fed back to them by removing the other option from the screen and moving the chosen option slightly to the bottom corner of the screen. This feedback simply showed the participants that their answer had been recorded. After each 50 trials, participants had the opportunity to take a short break, and they continued by pressing the space bar in a self-paced fashion. Experimental sessions lasted between 10 and 15 minutes depending on the average speed of responses.

3.1.4. Analysis

The data were analyzed using a linear mixed effect model using a binomial linking function with a geminate response (e.g., the word was perceived as *rikkeb*) being coded as 1 and a singleton response as 0. For the predictors, to limit the number of random effects and their correlations to be estimated, fixed effects were coded as numerical contrasts. Duration ranged from –2 to 2 in steps of 1, surrounding speech rate was coded as 0.5 for a fast rate and –0.5 for a slow rate. With this coding, an expected effect of speech rate and consonant duration should yield a positive regression weight, since geminate responses should be more frequent with longer consonant durations and, potentially, a faster speech rate.

For the three different continua, two independent linear contrasts were coded. Contrast coding allows a better control over the random effect structure in linear mixed effect models[6] and is potentially more powerful, as it eliminates the need for post-hoc tests

[6] Linear-mixed effect models with a binomial linking function often run into convergence problems that can be alleviated by using uncorrelated random effects. Correlation parameters cannot easily be removed for

requiring (e.g., Bonferroni) correction. The first contrast compared the fricative continuum to the two stop continua (*wasal-wassal*: –2/3, *rikeb-rikkeb* and *qata'-qatta'*: +1/3). A positive regression weight of this contrast would indicate more geminate responses for the stop than for the fricative continuum. Interactions would indicate that step and context-rate effects differ over continua. There is no a-priori reason to assume that it is the case, but they might indicate that rate effects differ between the continua. Stronger rate- and/or duration effects for the stop continua than for the fricative continuum would be reflected in a regression weight that has the same sign (i.e., both are either positive or negative) as the regression weight for main effect of step and/or context rate. Vice versa, if the simple effects of rate and/or duration are weaker for the stop continua than the fricative continuum, the regression weight for the interaction has the opposite sign than the main effect of step and/or context rate. The second contrast compared the two stop continua with each other (*wasal-wassal*: 0, *rikeb-rikkeb*: –1/2 and *qata'-qatta'*: +1/2). Note that this system of coding is an example of Helmert contrast coding (see, e.g., Field, Miles, & Field, 2012). Two-way interactions were specified with both the rate and the duration contrasts, and a random effect for participant with a maximal-random effect structure was specified (see also the note for **Table 1**).

3.2. Results

Figure 2 shows the mean proportion of geminate responses for each of the three continua and all combinations of consonant duration and ambient speech rate. The data show a clear rate effect for all three continua. This is reflected in the statistical analysis (see **Table 1**) which shows a strong rate effect, but with an interaction with the one of the continuum-type contrasts.

Therefore, separate analyses were run for each continuum with duration and rate as predictors. These showed a healthy rate effect for each continuum that was slightly smaller for the *qata'-qatta'* continuum (b = 1.696, SE(b) = 0.318, z = 5.343, $p < 0.001$) than for the other two continua (*rikeb-rikkeb*: b = 2.387, SE(b) = 0.360, z = 6.636, $p < 0.001$; *wasal-wassal*: b = 2.130, SE(b) = 0.324, z = 6.566, $p < 0.001$). Using the results of these analyses, we found the 50% point for both the slow and the fast continuum to estimate the magnitude of the boundary shift in milliseconds. A change in speech rate leads to a shift

Table 1: Results from the overall analysis of the likelihood of geminate responses, using the following specification: glmer (percQuantity ~ rateContrast * (isFric + betweenStops) + durContrast * (isFric + betweenStops) + (1 + rateContrast * (isFric + betweenStops) + durContrast * (isFric + betweenStops)||participant), family = binomial).

Predictor	b	SE(b)	z	*p*
(Intercept)	0.316	0.179	1.766	.077
Consonant Duration	3.024	0.222	13.595	<.001***
Speech Rate	2.125	0.272	7.799	<.001***
StopsvsFricatives	2.073	0.316	6.567	<.001***
betweenStops	–0.474	0.277	–1.712	0.087
Consonant Duration: StopsvsFricatives	0.240	0.142	1.690	.091
Consonant Duration: betweenStops	–0.440	0.179	–2.454	.014*
Speech Rate: StopsvsFricatives	0.099	0.239	0.413	.679
Speech Rate: betweenStops	–0.581	0.289	–2.013	.044*

categorical predictors with more than two levels in the function *glmer* from the package *lme4*, but this is possible for numerical predictors, which in turn allows a more conservative random effect structure.

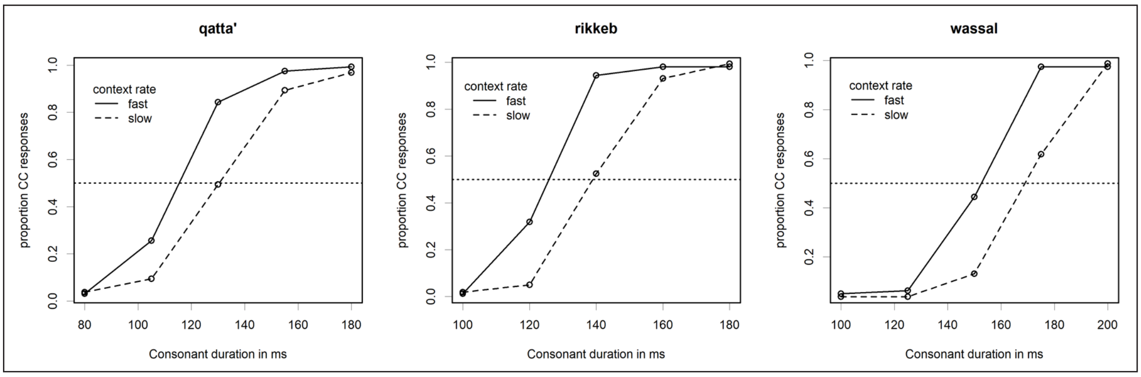

Figure 2: Average proportion of geminate responses for each of the three continua based on consonant duration and ambient speech rate. The horizontal line indicates the category boundary at 50%.

of 16 ms for *qata-‘qatta’* (114 vs 130 ms), a shift of 16 ms for *rikeb-rikkeb* (124 vs 140 ms), and a shift of 17 ms for *wassal* (151 vs 168 ms).

3.3. Discussion

This experiment asked the question whether ambient speech rate would influence the singleton-geminate distinction even when the durations of the immediately adjacent segments were kept constant. This is clearly the case; there were clear rate effects for all three continua. Not only the presence of such an effect is noteworthy but also its size. The category boundary is shifted by about 16 ms, which is about 11–12% of the mean boundary duration.[7] Despite not varying the duration of the surrounding vowels, this rate effect on the singleton-geminate distinction is larger than what has been observed for other contrasts even when the context directly preceding the critical phoneme was manipulated.

The data show that a quantity distinction can be influenced by distal speech rate. It may seem this simply replicates Hirata and Lambacher (2004), who found that surrounding speech rate—separated by one consonant from a target vowel that was either long or short—influences quantity perception in Japanese. The current data show that the same result can be obtained with two segments between target and context. At first sight, this may seem like a small difference, but there are two points to be considered here. First of all, a context is usually considered distal in the speech-rate normalization literature if it is at least a syllable away from the target (see, e.g., Heffner et al., 2017). This definition is supported given the importance of the syllable for rhythmically structuring language(s) (Ramus, Nespor, & Mehler, 1999). This is the case for the current study but not for Hirata and Lambacher (2004). Secondly, though the difference between one and two segments of a distance may seem small, it is after all an increase of 100% and as such sizeable.

This size of the speech rate effect found here strongly exceeds what Pind (1986) has found for Icelandic quantity (3 ms) for a manipulation of distal context only. This effect size is on par or even larger than what has been found for, for instance, vowel duration, in Dutch and German (Dutch: 10 ms = 8%, Reinisch & Sjerps, 2013; German: 10 ms = 7%, Reinisch, 2016). Although a direct comparison is difficult, because the studies differed in some respects (e.g., the amount of rate manipulation), such gross differences make it difficult to argue that the current shifts may be due to auditory processes. After all, auditory processes should not differ between Icelandic and Maltese listeners, and any

[7] To compare boundary size difference, we use proportional measures based on the Weber-Fechner law. Note however that this law is still an approximation and tends to overrate perceptual differences for small base quantities (e.g., the 1.1 ms difference for VOT found by Toscano and McMurray, 2015).

procedural differences are also unlikely to generate such strong differences. Importantly, this does not question that there are auditory processes contributing to rate normalization; it only suggests that they may be enhanced by language experience. Similar arguments have been made for vowel normalization by Sjerps et al. (2013) based on the finding that speaker normalization effects are larger the more speech-like the materials are.

Larger shifts for VOT are reported by Summerfield (1981) for initial stop voicing in English. He reports a shift from 20 to 27 ms (a shift of about 25% of the mean boundary size). However, in this stimulus set, there might be contribution of the perception of a prosodic boundary, which also leads to the expectation of an elongated VOT (Kim & Cho, 2013; Mitterer, Cho, & Kim, 2016). Indeed, Toscano and McMurray (2015) reported, for the same distinction, a much smaller shift of 1.1 ms (6%; this ratio may be an overestimation of the perceptual difference, given the issues of the Weber-Fechner Law with very small quantities).

4. General Discussion

This paper tested whether the singleton-geminate distinction in Maltese might profit from speech-rate normalization. Analyzing productions from Maltese speakers which produced sentences at different, but self-chosen rates showed that categorization accuracy based on duration improves when a speaker-specific boundary is used. These speaker-specific boundaries, in turn, are strongly correlated to a speaker's average speaking rate. Given that speech-rate normalization would hence be useful to improve categorization accuracy, a perception experiment tested whether listeners take the ambient speech rate into account when making singleton-geminate distinctions. To distinguish a speech-rate normalization account from an account based on vowel-consonant ratio as a rate-independent cue, only the distal context was manipulated in terms of speech rate. That is, the carrier phrase, but not the critical word containing the singleton/geminate contrast word medially, was manipulated in terms of speech rate. The results showed surprisingly strong effects of speech-rate normalization, even though the rate of the segments surrounding the critical singleton/geminate had a constant duration. These results indicate that the singleton-geminate distinction in Maltese should be considered rate dependent.

A first question that arises is why the production data for Maltese differ from those for Cypriot Greek (Arvaniti, 1999), where speech rate apparently does not affect the singleton-geminate boundary. First, this might be due to the procedural differences. In the study by Arvaniti (1999), participants were asked to read from cards, speaking either 'naturally' or fast. This method is problematic because it asks speakers to engage in meta-linguistic processing, they are reading from script, and speak faster as they would usually do. It is unclear whether effects found with such instructions would hold if variation is freely chosen by the participants. The rate differences in the Maltese data were due to tendencies by the different speakers, who generated sentences from a picture prompt, and hence did not read, but spoke at a self-chosen pace. These methodological differences are strong candidates to explain the difference in results.

Nevertheless, it is also possible that the languages differ in that respect. Effects of speech rate on temporal properties can vary over languages (Solé, 2007). The singleton-geminate difference has been shown to vary considerably across languages (Kingston et al., 2009; Yoshida et al., 2015). An interesting difference arises out of two recent studies that compared the consequences of mismatches for lexical access using priming (Kotzor, Wetterlin, Roberts, & Lahiri, 2016; Tagliapietra & McQueen, 2010). These two studies tested to what extent a mismatch in quantity can still lead to lexical access, that is, to what extent is a word with a geminate activated by input with a singleton (e.g., *rikeb* → *rikkeb*) and vice versa. These studies presented word fragments and estimated to what extent they lead to priming of target words. In Italian (Tagliapietra & McQueen, 2010), a mismatching prime with a singleton leads to stronger priming for a target containing a geminate than a mismatching prime with a geminate prime a target with a singleton. The opposite pattern is observed in Bengali (Kotzor et al., 2016), showing that it is not possible to easily generalize from one language to another.

A second question that arises is why the current context effects in perception are so strong, even surpassing the effects found in experiments in which distal and proximal context were rate manipulated. As Nakai and Scobbie (2016) noted, the utility of speech-rate normalization may differ per contrast (see also Port, 1979). For the English voiced/voiceless distinction, the data indicate that there is little need for normalization of VOT, as voiced stops do not have such strongly extended VOTs at slow rates so that they become like aspirated stops as produced in fast rates. Somewhat larger normalization effects have been observed for the Dutch /ɑ/-/aː/ distinction and the German distinction between /a/ and /aː/. The use of different IPA symbols for the Dutch but not the German contrast indicates the German but not the Dutch contrast relies solely on duration. Because Reinisch and Sjerps (2013) used a spectrally ambiguous vowel token, it may be argued that their data may overestimate the amount of rate normalization in real life. However, Bosker (2017) varied the spectral characteristics of the vowel and found that rate affects categorization independent of the spectral qualities. Indeed, the amount of rate normalization for the low vowel in Dutch and German seems similar when tested with similar methods (Reinisch, 2016; Reinisch & Sjerps, 2013). This pattern hence questions that the importance of rate normalization is proportional to the importance of duration for the distinction. If that were the case, rate normalization should be more effective for the German than for the Dutch contrast, yet the data (Reinisch, 2016; Reinisch & Sjerps, 2013) do not support this prediction. What might hence be crucial is how much overlap there is depending on rate. The Maltese singleton-geminate distinction then has a relatively strong overlap due to rate because, as the production data showed, both the singleton and the geminate category are affected by rate in production.

This is also relevant for the possibility mentioned in the introduction, that effects of distal rate seem stronger for segmentation than segment decisions. This was recently tested by Heffner, Newman, and Isardi (2017). They tested to what extent distal rate influences segmentation decisions (such as *Canadian notes* versus *Canadian oats*) and two types of segment decisions, word-initial voicing (e.g., *back* versus *pack*) and word-final voicing (e.g., *back* versus *bag*). Distal rate influences all decisions but those about word-initial voicing, hence showing that the dichotomy ‘segmentation versus segments’ does not hold. This is in line with the current data, which also show a strong dependency of ambient rate on a segmental decision. This suggests that what may matter is whether the durational cues change with rate in such a way that they endanger the category boundary. This may be the case for word-final stop voicing in English (which is mostly cued by vowel duration), but not for VOT for word-initial stop voicing (Nakai & Scobbie, 2016). Accordingly, rate normalization only is used for the former.

For the case of the quantity distinction in Maltese, the current data suggest that there is rate-dependence. The production study indicated that there is a strong overlap between the categories due to rate variation, which are not easily accounted for by just taking the duration of the neighboring vowels into account. In perception, listeners show strong differences in identification, which are much stronger than what has been found for VOT in English or quantity in Icelandic. This indicates that it is unlikely that the results are due to an auditory effect, and make it quite likely that the singleton-geminate distinction is rate-dependent in Maltese.

Additional File

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

- **Appendix.** Stimulus Material used in the production study (Mitterer, *Journal of Phonetics*, *66*, p. 28–44). DOI: https://doi.org/10.5334/labphon.66.s1

Acknowledgements

This work was supported by a University of Malta Research Grant to the author (CGSRP01-16).

Competing Interests

The author has no competing interests to declare.

References

Arvaniti, A. 1999. Effects of speaking rate on the timing of single and geminate sonorants. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, *1*, 599–602.

Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Bosker, H. R. 2017. Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*, *79*, 333–343. DOI: https://doi.org/10.3758/s13414-016-1206-4

de Jong, N. H., & Wempe, T. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. DOI: https://doi.org/10.3758/BRM.41.2.385

Dilley, L. C., & McAuley, J. D. 2008. Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294–311. DOI: https://doi.org/10.1016/j.jml.2008.06.006

Dilley, L. C., & Pitt, M. A. 2010. Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670. DOI: https://doi.org/10.1177/0956797610384743

Field, A., Miles, J., & Field, Z. 2012. *Discovering Statistics Using R* (1 edition). London; Thousand Oaks, Calif: SAGE Publications Ltd.

Galea, L. 2016. *Syllable structure and gemination in Maltese*. Universität Köln.

Hankamer, J., & Lahiri, A. 1988. The timing of geminate consonants. *Journal of Phonetics*, 16–327.

Heffner, C. C., Newman, R. S., & Idsardi, W. J. 2017. Support for context effects on segmentation and segments depends on the context. *Attention, Perception & Psychophysics*, *79*(3), 964–988. DOI: https://doi.org/10.3758/s13414-016-1274-5

Hirata, Y. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, *32*(4), 565–589. DOI: https://doi.org/10.1016/j.wocn.2004.02.004

Hirata, Y., & Lambacher, S. G. 2004. Role of word-external contexts in native speakers? Identification of vowel length in Japanese. *Phonetica*, *61*(4), 177–200. DOI: https://doi.org/10.1159/000084157

Idemaru, K., & Guion-Anderson, S. 2010. Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica*, *67*(1–2), 25–46. DOI: https://doi.org/10.1159/000319377

Kim, S., & Cho, T. 2013. Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America*, *134*(1), EL19–EL25. DOI: https://doi.org/10.1121/1.4807431

Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. 2009. Contextual effects on the perception of duration. *Journal of Phonetics*, *37*, 297–320. DOI: https://doi.org/10.1016/j.wocn.2009.03.007

Kotzor, S., Wetterlin, A., Roberts, A. C., & Lahiri, A. 2016. Processing of phonemic consonant length: Semantic and fragment priming evidence from Bengali. *Language and Speech*, *59*(1), 83–112. DOI: https://doi.org/10.1177/0023830915580189

Lubbers, M., & Torreira, F. 2013. *Praatalign: an interactive Praat plug-in for performing phonetic forced alignment*. Retrieved from: https://github.com/dopefishh/praatalign.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. 2008. Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071. DOI: https://doi.org/10.3758/PBR.15.6.1064

McMurray, B., & Jongman, A. 2011. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*, 219–246. DOI: https://doi.org/10.1037/a0022325

Miller, J. L., & Baer, T. 1983. Some effects of speaking rate on the production of /b/ and /w/. *The Journal of the Acoustical Society of America*, *73*(5), 1751–1755. DOI: https://doi.org/10.1121/1.389399

Miller, J. L., Green, K. P., & Reeves, A. 1986. Speaking Rate and Segments: A Look at the Relation between Speech Production and Speech Perception for the Voicing Contrast. *Phonetica*, *43*(1–3), 106–115. DOI: https://doi.org/10.1159/000261764

Mitterer, H. 2018. Not all geminates are created equal: Evidence from Maltese glottal consonants. *Journal of Phonetics*, *66*, 28–44. DOI: https://doi.org/10.1016/j.wocn.2017.09.003

Mitterer, H., Cho, T., & Kim, S. 2016. How does prosody influence speech categorization? *Journal of Phonetics*, *54*, 68–79. DOI: https://doi.org/10.1016/j.wocn.2015.09.002

Mitterer, H., & Reinisch, E. 2013. No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*, 527–545. DOI: https://doi.org/10.1016/j.jml.2013.07.002

Nakai, S., & Scobbie, J. 2016. The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1). DOI: https://doi.org/10.5334/labphon.49

Newman, R. S., & Sawusch, J. R. 1996. Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics*, *58*, 540–560. DOI: https://doi.org/10.3758/BF03213089

Newman, R. S., & Sawusch, J. R. 2009. Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics*, *37*(1), 46–65. DOI: https://doi.org/10.1016/j.wocn.2008.09.001

Peirce, J. W. 2007. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. DOI: https://doi.org/10.1016/j.jneumeth.2006.11.017

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. DOI: https://doi.org/10.1159/000028448

Pickett, J. M., & Decker, L. R. 1960. Time factors in perception of a double consonant. *Language and Speech*, *3*(1), 11–17. DOI: https://doi.org/10.1177/002383096000300103

Pind, J. 1986. The perception of quantity in Icelandic. *Phonetica*, *43*(1–3), 116–139. DOI: https://doi.org/10.1159/000261765

Pind, J. 1995. Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics*, *57*(3), 291–304. DOI: https://doi.org/10.3758/BF03213055

Port, R. F. 1979. The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics*, *7*(1), 45–56.

Ramus, F., Nespor, M., & Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. DOI: https://doi.org/10.1016/S0010-0277(99)00058-X

Reinisch, E. 2016. Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics*, *37*(06), 1397–1415. DOI: https://doi.org/10.1017/S0142716415000612

Reinisch, E., Jesse, A., & McQueen, J. M. 2011. Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 978–996. DOI: https://doi.org/10.1037/a0021923

Reinisch, E., & Sjerps, M. J. 2013. The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101–116. DOI: https://doi.org/10.1016/j.wocn.2013.01.002

Sjerps, M. J., McQueen, J. M., & Mitterer, H. 2013. Evidence for precategorical extrinsic vowel normalization. *Attention, Perception, & Psychophysics*. DOI: https://doi.org/10.3758/s13414-012-0408-7

Solé, M.-J. 2007. Controlled and mechanical properties in speech. In: Solé, M.-J., Beddor, P. S., & Ohala, M. (eds.), *Experimental approaches to phonology*, 302–321. Oxford: Oxford University Press.

Strunk, J., Schiel, F., & Seifart, F. 2014. Untrained forced alignment of transcriptions and audio for language documentation corpora using WebMAUS. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Piperidis, S., et al. (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26–31, 2014*, 3940–3947. European Language Resources Association (ELRA). Retrieved from: http://www.lrec-conf.org/proceedings/lrec2014/summaries/1176.html.

Summerfield, Q. 1981. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1074. DOI: https://doi.org/10.1037/0096-1523.7.5.1074

Tagliapietra, L., & McQueen, J. M. 2010. What and where in speech recognition: Geminates and singletons in spoken Italian. *Journal of Memory and Language*, *63*(3), 306–323. DOI: https://doi.org/10.1016/j.jml.2010.05.001

Toscano, J. C., & McMurray, B. 2012. Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics*, *74*(6), 1284–1301. DOI: https://doi.org/10.3758/s13414-012-0306-z

Toscano, J. C., & McMurray, B. 2015. The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience*, *30*(5), 529–543. DOI: https://doi.org/10.1080/23273798.2014.946427

Yoshida, K., de Jong, K. J., Kruschke, J. K., & Päiviö, P.-M. 2015. Cross-language similarity and difference in quantity categorization of Finnish and Japanese. *Journal of Phonetics*, *50*, 81–98. DOI: https://doi.org/10.1016/j.wocn.2014.12.006

How to cite this article: Mitterer, H. 2018 The singleton-geminate distinction can be rate dependent: Evidence from Maltese. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9(1): 6, pp. 1–16, DOI: https://doi.org/10.5334/labphon.66

Submitted: 12 December 2016 **Accepted:** 04 February 2018 **Published:** 27 April 2018

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

OPEN ACCESS