

## Research Article

## Not all geminates are created equal: Evidence from Maltese glottal consonants

Holger Mitterer

*Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida MSD 2080, Malta*

## ARTICLE INFO

*Article history:*

Received 3 October 2016

Received in revised form 14 September 2017

Accepted 15 September 2017

*Keywords:*

Geminates

Perception

Production

Features

Maltese

## ABSTRACT

Many languages distinguish short and long consonants or singletons and geminates. At a phonetic level, research has established that duration is the main cue to such distinctions but that other, sometimes language-specific, cues contribute to the distinction as well. Different proposals for representing geminates share one assumption: The difference between a singleton and a geminate is relatively uniform for all consonants in a given language. In this paper, Maltese glottal consonants are shown to challenge this view. In production, secondary cues, such as the amount of voicing during closure and the spectral properties of frication noises, are stronger for glottal consonants than for oral ones, and, in perception, the role of secondary cues and duration also varies across consonants. Contrary to the assumption that gemination is a uniform process in a given language, the results show that the relative role of secondary cues and duration may differ across consonants and that gemination may involve language-specific phonetic knowledge that is specific to each consonant. These results question the idea that lexical access in speech processing can be achieved through features.

## 1. Introduction

Many languages distinguish short and long consonants and those languages are not necessarily related (e.g., Finnish, Italian, Japanese, and Maltese). It is generally accepted that long consonants, or geminates, are longer than short consonants, or singletons. This difference in duration is the primary distinction made in production and used in perception (for a recent overview, see [Table 1 in Hamzah, Fletcher, & Hajek, 2016](#)). However, there is considerable discussion regarding the status and origin of secondary cues that are non-durational (e.g., release burst amplitude in stops, see [Hamzah, Fletcher, & Hajek, 2012](#)). The purpose of the current paper is to show that secondary cues may be stronger or weaker depending on which segment is geminated. As a consequence, geminates may sometimes be viewed as segments in their own right that have articulatory and acoustic properties that are not always easily derived as some form of strengthening of the singleton.

In previous research on geminates, there is some disagreement on the importance of secondary cues. Some argue that segment duration really is the main cue and that other cues may be perceptually close to irrelevant ([Kotzor, Wetterlin,](#)

[Roberts, & Lahiri, 2016](#)), but others argue that secondary cues may be part and parcel of the distinction ([Yoshida, de Jong, Kruschke, & Päiviö, 2015](#)). [Yoshida et al. \(2015\)](#) varied the duration of the closure of a singleton /p/ or geminate /p:/ and tested the relative contribution of closure duration and the base (i.e., whether the stimulus was originally a singleton or a geminate). To do so, they asked Finnish and Japanese participants—both familiar with singleton-geminate contrasts in their native language—whether they perceived a singleton or a geminate. Both Finnish and Japanese stimuli were presented to Finnish and Japanese listeners. The results showed strong effects of the base, sometimes with differences of about 40% in quantity categorization. That is, a closure duration of 105 ms would be perceived as geminate in only 10% of the cases when the base stimulus originally contained a singleton, but as a geminate in 50% of the cases when the base stimulus originally contained a geminate. Interestingly, they found that such effects were, to some extent, language independent. That is, similar effects were found when Finnish listeners categorized Finnish and Japanese stimuli (and vice versa), indicating that some secondary cues are similar across languages.

[Yoshida et al. \(2015\)](#) explain this result by arguing that the quantity contrasts change the word prosody. This fits nicely with other proposals that the main difference between

**Table 1**

Outcome of the linear mixed-effect model predicting the amount of voicing leak depending on the underlying segment and its quantity.

Term	Estimate (ms)	<i>t</i>	<i>p</i>
(Intercept)	35.75	22.89	<0.001
Quantity	−9.94	−5.66	<0.001
Glottal vs Oral	20.21	5.43	<0.001
/s/ vs oral stop	−5.15	−1.42	0.16
/h/ vs glottal stop	−43.33	−5.65	<0.001
Quantity x Glottal vs Oral	−15.18	−3.64	<0.001
Quantity x /s/ vs oral stop	0.05	0.01	0.99
Quantity x /h/ vs glottal stop	5.31	0.75	0.46

singletons and geminates is rhythmic/prosodic. [Ridouane \(2010\)](#) argued that geminates are associated with two timing slots and are additionally supplied with the feature *tense*, leading to a more forceful articulation. Evidence for this claim stems from acoustic and articulatory measures that show that geminates are produced more forcefully, for instance, with larger alveolar tongue contact for geminate /t:/ than singleton /t/. [Kotzor et al. \(2016\)](#) proposed that the primary distinction between singleton and geminates is that words with the geminate contain an extra mora (following [Hayes, 1989](#)) and that such a difference in rhythmic properties is necessary and sufficient for speakers to implement the distinction and for listeners to hear the distinction. That is, a geminate segment can be decomposed into the properties of the respective singleton segment plus some gemination property even in its phonetic implementation. Even though such accounts may seem rather different, it is more difficult to distinguish between them than it appears. The assumption that a geminate is connected to an additional mora means that the geminate would have more prosodic weight. This, in turn, would mean that geminates are a strengthened articulation in comparison to the singleton, also realized as increased duration ([Cho & McQueen, 2005](#); [Cho, McQueen, & Cox, 2007](#)). Both assumptions—an extra timing slot or an extra mora for a geminate—hence predict that geminates are longer and have a strengthened articulation in comparison to the singleton.

This would mean that there are differences between singleton and geminates other than the duration of the segment itself and that those differences may be language-independent. While [Yoshida et al. \(2015\)](#) found evidence for such language-independent secondary cues for gemination, there is also evidence that there are language-specific ways to enhance the singleton-geminate distinction. One focus of previous research was the duration of the neighbouring vowels. Most languages show a contrastive pattern, so that vowels are phonetically shorter next to long consonants. Japanese provides a counterexample to this pattern by lengthening vowels before geminates ([Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009](#)). These effects lead to language-specific learning, so that listeners will use vowel length to categorize consonants as singletons or geminates according to the production pattern in their native language. Japanese listeners give more geminate responses when the preceding vowel is long, while Norwegian and Italian listeners—from languages with a contrastive pattern—give fewer geminate responses when the preceding vowel is long (see [Kingston](#)

[et al., 2009](#)). This suggests that there also is language-specific phonetic knowledge ([Kingston & Diehl, 1994](#)) on how to implement the singleton-geminate contrast.

Despite such differences in views regarding secondary cues to gemination, there is one common assumption to these approaches ([Kotzor et al., 2016](#); [Yoshida et al., 2015](#)): That gemination works similarly for all consonants that are geminated. While there may be secondary cues which are specific to some segments (e.g., burst amplitude which only occurs in stops, but not in nasals, since they do not have a burst), it is nevertheless assumed that, underlyingly, the planning and execution of singletons and geminates is governed by the same principles, be it a stronger articulation due to a feature [TENSE] or more prosodic weight due to an additional mora ([Kotzor et al., 2016](#); [Ridouane, 2010](#)). This is in line with the prevalence of feature theories in linguistics ([Embick & Poeppel, 2015](#)), which assume that segments are not primary in speech processing, but that the features (and timing properties) that define these segments are the primary objects of speech perception. The objective of this paper is to question such one-size-fits all approaches to gemination.

There is, for instance, an informal observation about the singleton-geminate distinction that challenges this general approach. [Ladefoged and Maddieson \(1996, p. 75\)](#) noted that the glottal stop is by default not really a stop but tends to surface as glottalization with no stop closure. A stop-like pronunciation is only reliably observed when the glottal stop occurs as a geminate. This would suggest that the acoustic cues for the singleton-geminate distinction would vary over place of articulation of a stop. For oral stops, the main cue would be constriction duration, while there is an additional difference for glottal stops, so that the singleton glottal stop surfaces as a glottalization while the geminate glottal stop surfaces as a stop.

Even if this is borne out by more thorough empirical investigation—which is one goal of the present paper—this may still be explained, though not very well, as an example of strengthening. After all, oral stops may also undergo lenition in connected speech and surface as flaps ([Warner & Tucker, 2011](#)). [Ridouane \(2010\)](#) argued that gemination leads to protection from lenition, so that singleton but not geminate stops can be produced without a release burst. The pattern observed by [Ladefoged and Maddieson \(1996\)](#) may hence be viewed in terms of (prosodic) strengthening of the glottal stop by assuming that the effects of the strengthening affect the acoustic outcome in a non-linear fashion. Even with such an explanation, such a finding would still raise the question whether features can be primary in speech processing. After all, the listener only has access to the auditory speech signal and when the acoustic consequences of gemination—whatever feature or timing unit it is associated with—differ radically between segments, it is questionable whether it would be functional to assume that the listener makes use of such features in speech perception. The critical question here then is whether the implementation of gemination differs strongly between oral and glottal stops and, if that is the case, whether the secondary cues that are specific to the glottal stop are important in perception. Answering this question is the first aim of this paper.

As will be reviewed below, Maltese is a language in which minimal pairs differing only in consonant quantity can be elicited for both oral and glottal stops, thanks to the derivational

morphology of verbs. In Maltese (see Galea, 2016), geminates can occur in all syllable positions and all consonants can be geminated but not all combinations thereof are possible. The glottal stop [ʔ] and fricative [h] do not occur as initial geminates. Word-initial geminates, moreover, tend to trigger an epenthetic vowel (e.g., *pparkja*, Engl., ‘he parked’, tends to be produced as [ip:erkjɛ]), so their word-initial status is somewhat questionable. Moreover, Galea (2016) found that Maltese listeners cannot perceive the singleton-geminate contrast in word-initial position when the epenthetic vowel is spliced off. Surprisingly, this was also the case for fricatives for which segment duration is audible in utterance initial position. Word-final contrasts are enhanced by a phonological vowel-duration contrast, so it is not surprising that the vowel before a geminate is phonetically shorter (Hume, Rose, & Spagnol, 2014). The current paper will focus on word-medial geminates for two reasons. First, glottal obstruents do not have a quantity contrast in word-initial position and, second, because word-medial minimal pairs are frequent due to the derivational morphology of Semitic verbs (see below for details).

As it turns out, the second glottal phoneme of Maltese, the fricative /h/, provides another interesting case to investigate secondary cues in geminates. In Germanic languages, /h/ is often restricted to the syllable-onset position (e.g., Wiese, 1996). This is often explained by the low amplitude of /h/, which makes it difficult to transmit in coda position (Rietveld & Van Heuven, 1997). In Maltese, /h/ can occur in onset and coda position not only as a one-segment coda or onset but also in clusters (e.g., *bħala*, /bħala/, Engl. ‘like’, *qamħ*, /ʔamħ/, Engl., ‘wheat’). If it occurs in coda position, it is quite often produced with a secondary constriction. Azzopardi-Alexander & Borg, 1997 had noted that /h/ can also be produced as a voiceless pharyngeal fricative, varying over speaker and position. Anecdotally, there seems to be a range of possible pronunciations, facilitated by the fact the fricative inventory in Maltese has a large gap between the post-alveolar /ʃ/ and /h/. This large gap is then filled with various allophones for /h/. Fig. 1 provides examples coming from the same speaker, in which the /h/ is produced once as fully voiced—a production frequently found in Germanic languages (e.g., for English, see Pierrehumbert & Talkin, 1992)—and once as an unvoiced fricative with an additional oral constriction, which is here transcribed as velar. A full description of this allophony is beyond the scope of this paper, but impressionistically, a second constriction is used when /h/ occurs in coda position and especially after back vowels. Moreover, Maltese speakers sometimes assimilate Dutch [x] to their native /h/.

For the current purposes, it suffices to say that /h/ occurs in two different forms in Maltese: either as a “weak” glottal fricative, which is often realized as voiced, or as a “strong” voiceless fricative with an additional oral constriction, which may vary from pharyngeal to velar. One can therefore ask the question whether the “strong” form of /h/ is more likely to be used for the geminate than for the singleton. Importantly, this additional constriction of the strong form cannot be construed as a strengthening of the original laryngeal gesture for /h/. It would hence require the assumption that geminate /h:/ in Maltese is a segment that has “unique cues”. “Unique” here means that these cues cannot be easily understood as some form of strengthening of the articulatory gestures of the singleton and

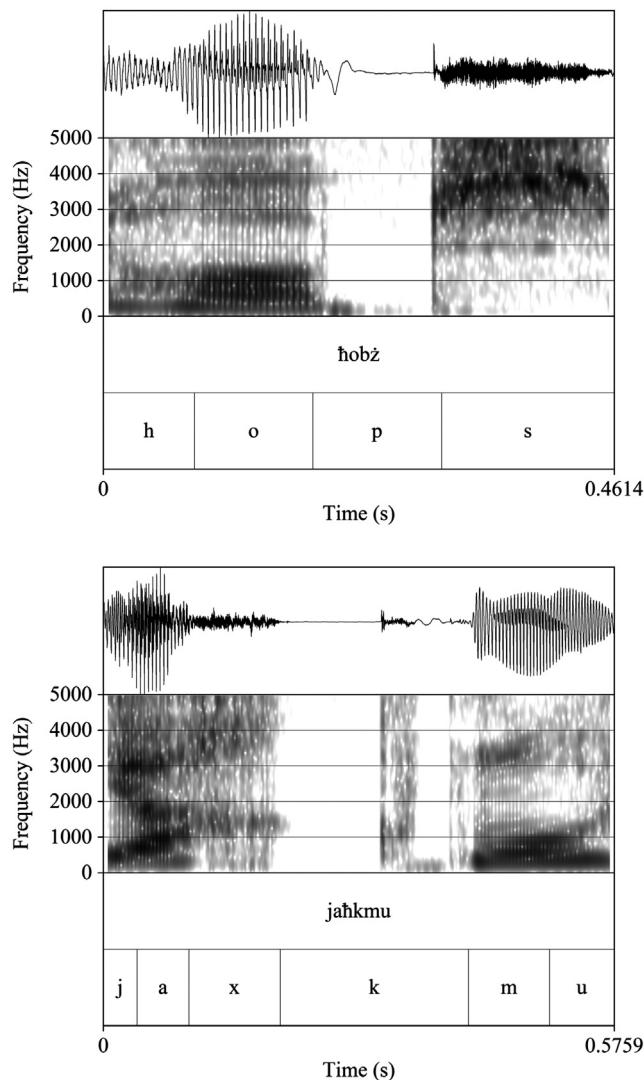


Fig. 1. Two examples of /h/ in Maltese produced by the same speaker in connected speech. In the upper panel, /h/ is realized as a breathy vowel in *ħobż* /ħobz/ (Engl., ‘bread’); in the lower panel, /h/ is produced as (probably) a velar fricative in *jahkmu* /jahkmu/ (Engl., ‘they fight’). [Multimedia files available in the [Supplementary data](#) for this article].

are specific to the geminated consonant in question. Consequently, the geminate /h:/ may need to be considered as a segment in its own right in both perception and production, rather than a tense or prosodically strong variant of /h/.

If we observe that an additional gesture is used for geminate /h:/, it provides a counterexample to the idea that geminates are generated by strengthening and lengthening the articulation of the singleton. If we can further show that such additional cues are specific to /h/, it would provide an example that, at the phonetic implementation level, segments are in fact atomic and cannot be decomposed easily into features during speech processing, because feature combinations bring forth irreducible properties. Similarly, this would apply to the glottal stop, which would surface in an acoustically distinct form as singleton (as glottalization) and geminate (as a stop).

To investigate these issues, Experiment 1 elicited productions of glottal and—as a control “group”—oral segments as singletons and geminates in minimal pairs. Experiment 2 then

builds on the findings of the production experiment and investigates the importance of durational and non-durational cues for the singleton-geminate distinction for oral and glottal segments.

## 2. Experiment 1: Production

Maltese is often described as a mixed language of Semitic origin with strong influences from Italian and English. This is evident in the verb system, in which verbs of Semitic and other origins differ in their derivational and inflectional morphology. Imported verbs from Italian (e.g., *(i)kkanta*, Engl., *to sing*) and English (e.g., *(i)pparkja*, Engl., *to park*) are affixed to indicate tense, person, and number (e.g., *nipparkja*, Engl., *I park*, *ipparkjat*, Engl., *she parked*). Semitic verbs are based on tri-consonantal roots (e.g., *k-t-b*, Engl., *to write*) and take part in an extensive inflectional and derivational morphology, in which it is not uncommon that a single root gives rise to several hundred surface forms.

We make use of this system by eliciting the 3rd male singular past tense form in the first form (e.g., *w-q-f*, Engl., *to stop* → *waqaf*, Engl., ‘he stopped’, in the sense of stop walking, running, etc.) and in the second form, which indicates a causative meaning (e.g., *waqqaf*, Engl., ‘he stopped’, in the sense of stop something/someone else). The second form is derived by making the second root consonant (i.e., the /ʔ/ in *waqaf* /waʔaf/) a geminate (i.e., *waqqaf* /waʔ:af/) These forms provide minimal pairs that only differ in the quantity of the second root consonant. The 3rd male singular forms were chosen for elicitation because these are the only forms in which the first and second “binyam” form of a verb constitute a minimal pair with an intervocalic singleton-geminate distinction. Note that geminates do not occur only in second-form verbs, but also, for example, in nouns (e.g., *qattus*, /ʔat:us/, Engl., *cat*), adjectives and adverbs (e.g., *ezatt* /ezat:/, Engl., *exact(ly)*).

We elicited such minimal pairs arising from verbs with glottal stop, an oral stop (/t/ or /k/), an /h/, or an /s/ as middle root consonant. The oral stops and the fricative /s/ serve as control conditions to test whether the effects of gemination differ between glottal and oral consonants. The forms were elicited using a sentence guessing task to avoid a reading task (see below for details).

### 2.1. Method

#### 2.1.1. Participants

Fourteen native speakers of Maltese (eight female/six male) participated in the production task. They were students at the University of Malta aged 19–26. Participants first signed a consent form that informed them that their utterances were recorded. They were also asked whether they would agree that their utterances would be made publicly available, which all of them agreed to. The procedure has been cleared with the University of Malta ethics review board (Project: Spoken-word recognition in Maltese). Participants were paid for their participation.

#### 2.1.2. Materials

Starting with a list of all Semitic roots used in Maltese, we identified as many roots as possible for each target consonant

for which both the first and second form is commonly used in Maltese. This resulted in eight each for /h/ and /ʔ/, nine for /s/, and no more than six for any of the oral stop consonants. From this set, all the roots for /h/, /ʔ/, and /s/ were used, and then a fourth generic category of “oral stop” was created with six roots with /t/ and two with /k/ as middle root consonants. For these verbs, sentences were generated in which the verb occurred in either the first or the second form for the 3rd male singular person. The sentences all had a common form, such as *Yesterday, Matthew slept on the sofa*. That is, the sentences all started with a time indication to induce use of the past tense (e.g., *yesterday*) and then continued as SVO sentences (all verbs and sentences including an English translation are provided in the [Appendix A](#)).

Next, visual prompts were created that sought to elicit these sentences in a sentence-generation task. [Fig. 2](#) provides two examples of such displays, which consisted of a written time indication (e.g., *il-lejl li għadda*, Engl., *last night*, literally, *the night that passed*), an actor on the left, a verb root in the middle, and an object on the right. Participants were instructed to generate a sentence out of these clues starting with the time indication, then using the character or characters, the root and the object on the right as subject, verb, and object in that order. The participants were familiarized with four actors before the experiment, and hence knew that the actor in [Fig. 2](#) went by the name of *Matthew*. To help participants remember the name of the actors, all had their initial on their sweater or t-shirt. To prompt them to start with the time indication, the arrow and the text *ibda hawn* (Engl., ‘start here’) appeared next to this time indication. Pilot testing had shown that this was necessary to ensure the sentences would start with the time indication. [Fig. 2](#) provides examples of such prompts, including the intended sentence. There were hence 66 sentences to elicit, each with a different prompt. As fillers,<sup>1</sup> participants also generated 66 verbs in the present tense, using the 3rd plural form by presenting two actors (e.g., *Matthew and Daniel*).

#### 2.1.3. Procedure

After filling in an informed-consent form, participants were first familiarized with the task using an example display on paper and written instructions. Next, participants were familiarized with the four characters and their names (e.g., the character called *Matthew* in [Fig. 2](#)). Then they performed the sentence generation task, which was run with the programme *SpeechRecorder* ([Draxler & Jänsch, 2004](#)) and conducted in a sound-proof booth. To increase the amount of data, each prompt was presented twice in sets of five. That is, five prompts were selected for a block of ten trials. During the first five trials from each block of ten, each of the five prompts was presented once with the instruction “Can you guess this sentence?”. This was followed by a second set of five trials using the same five prompts again but now in a different order, with the instruction “Can you remember this sentence?”.

In *SpeechRecorder*, participants see a cartoon version of a traffic light on the screen which goes from red to amber to green and back to red. Participants were instructed to prepare

<sup>1</sup> These forms are fillers for the present purpose, but the 3rd plural present tense form contains the three root consonants in sequence (e.g., *jiktbu*, Engl., *they write*). These forms were elicited to see how faithfully these complex clusters would be produced. They also served the purpose to obfuscate our interest in eliciting minimal singleton-geminate pairs.

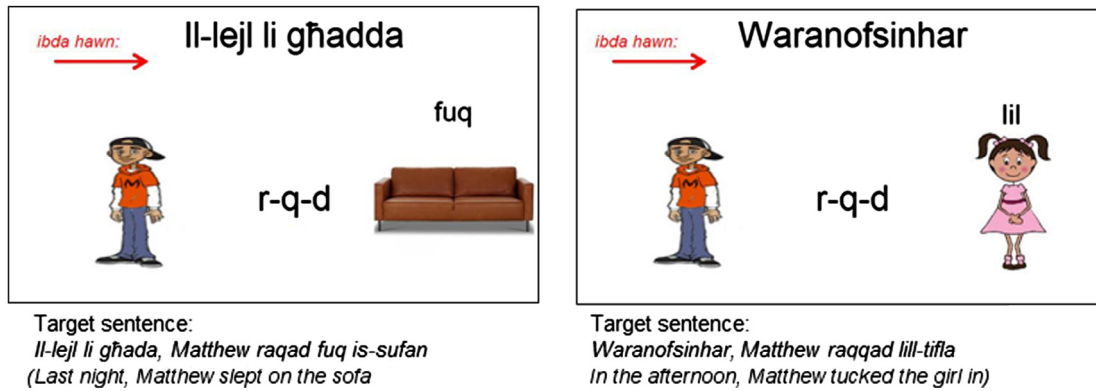


Fig. 2. Sentence prompts as used to elicit Semitic verbs in the first and second form (using 3rd male singular in the past tense), which form minimal pairs differing in singleton versus geminates.

the sentence while the traffic light was red (which lasted 4 s when a prompt was seen for the first time and 2 s when seen for the second time). They were instructed to start speaking only when the traffic light went green. If the participant produced the intended verb form during the recording time of 4 s, the next trial was initiated, if not, this trial was repeated. If the participant did not produce the correct form on the third attempt, the trial was abandoned.

The recording session was split into two halves, each half taking about 20 min. Between the two sessions, participants were invited to leave the recording booth and have some water. Minimal pairs were divided over blocks, that is, if a participant had the target form *raqqad* in the first block, the target form *raqqad* occurred in the second block. To balance how often each form occurred in which block, two different lists were prepared, and forms that occurred in the first block on the first list occurred on the second block on the second list (and vice versa).

#### 2.1.4. Data analysis

To analyse the production data, the occurrence of the verb form was first marked by hand in a Praat (Boersma, 2001) TextGrid Object. A native speaker of Maltese coded if the target form was produced correctly or not and if so, indicated when during the recording the target form occurred. Next, these forms were then subjected to forced alignment using the language-independent phone set from the Munich Automated Segmentation (MAUS) system (Strunk, Schiel, & Seifart, 2014) and Praatalign (Lubbers & Torreira, 2013). The Maus system has been shown to be highly accurate with an over 97% agreement with highly trained phoneticians (Schiel, 1999). Forced-alignment is a well-established tool, with proven reliability that even has been successfully used to analyse spontaneous speech with much more phonetic variation (Schuppler, Ernestus, Scharenborg, & Boves, 2011). For small languages, like Maltese, the MAUS system uses the acoustic models with the most training data from all languages it has been trained on. For the critical segments used in this study (/t/, /k/, /s/, /h/, /ʔ/, and /x/), these are the German phone models but for the initial /w/, for instance, an English phone model is used.

Praatalign uses the MAUS acoustic models and allows the alignment of given words with variants as supplied in a cus-

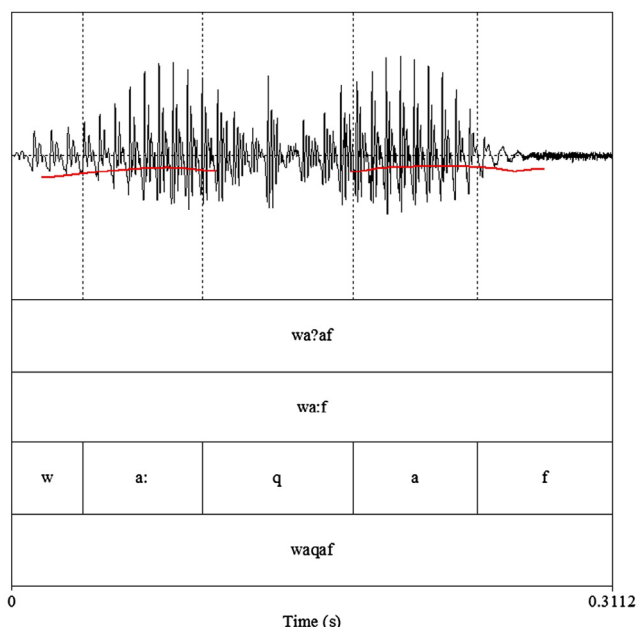
tomized lexicon. This allows us to measure the duration of the different segments (and to some extent how they were produced) in a reliable and replicable way. To measure whether /h/ was produced with an additional oral constriction, the forced-alignment was given the option that words with a medial /h/ (such as *wahaʔ* and *wahhaʔ*) were produced either with [h] or with the velar [x]. Velar [x] was used rather than another back fricative because [x] is articulatorily and acoustically more distinct from /h/ than other more back fricatives, such as a pharyngeal fricative. If the forced-alignment algorithm prefers [x] over [h], this is a strong indication that there is a secondary oral constriction.

For items with stops, a form with deletion of the stop was also possible. Consequently, when the glottal stop was only realized as glottalization, the forced-alignment algorithm preferred the version with no glottal stop and a long vowel over a form with a glottal stop. Fig. 3 provides an example of such a case. The algorithm prefers the transcription [wa:f] over [waʔaf]. For these forms, the duration of glottalization was estimated by the author based on the drop or absence of  $f_0$ . These were then marked as “q” in the TextGrid (see Fig. 3). A subset of those ( $n = 40$ ) were also rated by a second rater, who was a trained Maltese phonetician. The correlation between the duration ratings was 0.86, showing that these hand-coded durations were relatively reliable. For glottal stops, the realization (glottal stop vs glottalization) was hence coded based on whether the forced-alignment algorithm found a glottal stop or not. (As for the coding of the realization of /h/, this criterion is objective and replicable.)

After forced alignment, the following measures were taken: For the stops and the fricatives, the amount of voicing during the medial consonant was measured. For the alveolar fricatives, the centre of gravity during the middle 10 ms of frication was measured to estimate whether the constriction location differs between singleton /s/ and geminate /s:/. (Note that for /h/, the constriction location was estimated by using variants in the forced alignment.)

## 2.2. Results

From the 1750 recordings, 331 (17.7%) were rejected because they did not contain the intended target form. Three of the remaining 1439 utterances (two singleton glottal stops



**Fig. 3.** Example of the forced alignment for a glottal stop. The first tier shows the canonical form, the second the transcribed form, indicating that no glottal stop was found. The third shows the inserted glottalization period based on the pitch (red/grey line) and the fourth tier the orthographic form. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and one singleton /t/) were aligned with a deleted stop and excluded from further analysis. This left 1436 data points for the analysis.

### 2.2.1. Duration

Unsurprisingly, there were strong duration differences between singletons and geminates, with geminates being 1.6 (for /s/) to 1.9 (for /h/) times as long as the singletons (see Fig. 4). Maybe even more important, the effect sizes indicate how well the two categories are separated by duration alone. As it turns out, the oral segments were slightly better separated (oral stops: 2.28 and for /s/: 2.41, effect size measure Cohen's  $d$ ) than the glottal segments (glottal stop: 1.63, /h/: 1.94).

### 2.2.2. Secondary cues: Realization

As the different measures of segment realization are incommensurable between segments, segment realization was investigated for each segment independently. For the glottal stop, it was analysed whether the segment realization (full glottal stop or glottalization) depended on the segment quantity. The left panel of Fig. 5 shows that this was clearly the case. Most of the singleton glottal stops were not recognized as such by the forced alignment system since they were mainly realized as glottalization without discernible closure (in line with the informal observation by Ladefoged & Maddieson, 1996). In contrast, most geminate glottal stops were realized as glottal stops that contained a clear closure. This pattern was also, unsurprisingly, statistically significant. This was tested with a generalized linear-mixed effect model with Quantity (singleton vs geminate) as the independent variable and segment realization (glottal stop vs glottalization) as the dependent variable. Item and Participant were added as random effects with a maximal-random effect structure (Barr, Levy, Scheepers, &

Tily, 2013). To account for the categorical nature of the dependent variable, a binomial linking function was used. The results provided an intercept of 2.30 logOdds. This represents the estimated likelihood for a glottalized realization for singletons, which were mapped on the intercept. The analysis also provided a significant beta weight for quantity ( $b = -6.07$ ,  $z = -3.54$ ,  $p < .001$ ) that indicated a lower likelihood for a glottalization for geminates.

For the oral stops, the forced-alignment found a stop in all but one case, where the stop seemed to be deleted. This indicates that both singleton and geminate oral stops were realized as such in Maltese without reduction. That is, different phonetic gestures were used for singleton versus geminate glottal stops, but oral stops did not strongly differ in their implementation between singleton and geminates.

A strong difference in segment realization between singleton and geminates was also found for /h/, which was predominantly aligned as [h] when a singleton but as [x] when a geminate. This was also tested with a linear-mixed effect model with quantity (singleton vs geminate) as the independent variable and segment realization ([h] or [x]) as the dependent variable. Item and Participant were added as random effects with a maximal-random effect structure (Barr et al., 2013). To account for the categorical nature of the dependent variable, a binomial linking function was used. The analysis provided an intercept of 2.05, which is the logOdds value for the likelihood of a [h] transcription for singletons, and this likelihood is significantly lower for geminates ( $b = -4.38$ ,  $z = -7.88$ ,  $p < .001$ ).

The data for /h/ indicate that there is a different place of articulation for the geminate than for the singleton. To test whether there is a similar difference for /s/, we measured the spectral centre of gravity of /s/ at the midpoint of frication and tested whether this measure (which is a gradient measure of place of articulation) differed between singleton and geminates. One could expect that the /s/ might be "sharper" for the geminate than singleton due to strengthening, leading to a higher spectral centre of gravity. The small difference in this direction was marginally significant (a linear mixed-effect model with subject and item as random effects and a maximal random effect structure estimates a difference of 388 Hz,  $t = 2.07$ ,  $p = .058$ , with an intercept of 7253 Hz for the singletons). Maybe more importantly, the effect size is small to moderate ( $d = 0.43$ , with a residual standard deviation of 894 Hz). While we find a very clear difference for /h/ in terms of place of articulation, only a small effect is observed for /s/.

Another way to test the role of such secondary cues is to test whether they help to predict phonological quantity better once duration is already taken into account. Surprisingly, this was not the case: We ran linear mixed-effect models with a binomial linking function predicting segment quantity based on duration alone or based on duration and segment realization. For both /h/ and glottal stop, the model comparison showed no difference ( $\chi^2 < 1$ ), hence indicating that taking the secondary cues into account does not lead to a better prediction. However, the reverse comparison showed the same (i.e., duration of the segment does not allow a better prediction if the realization is already taken into account, both  $\chi^2 < 1$ ). The strong collinearity between the two factors made it difficult to estimate their independent contribution. For /s/, adding the

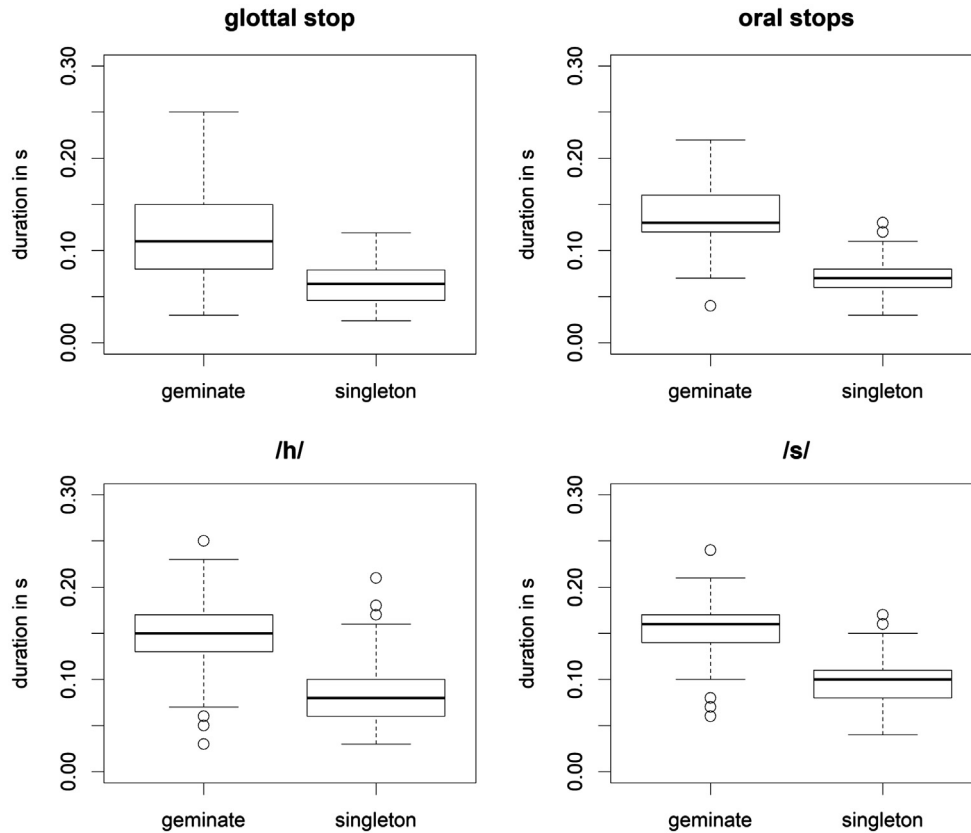


Fig. 4. Boxplot of the durations of singletons and geminates for the four types of segments.

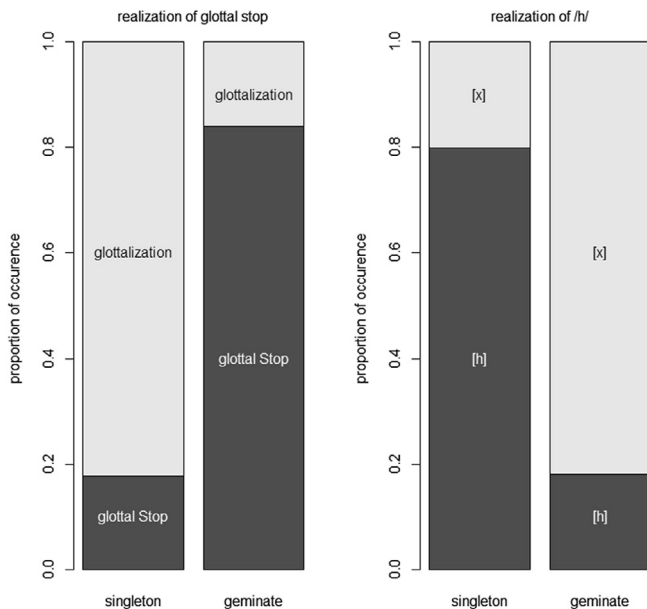


Fig. 5. Realization of the glottal stop and /h/ depending on segment quantity.

secondary cue CoG above duration does not improve model fit ( $\chi^2 < 1$ ), but adding duration to a model with just the secondary cue does ( $\chi^2(1) = 53.45$ ,  $p < 0.001$ ).

### 2.2.3. Secondary cues: Voicing during constriction

Another potential secondary cue distinguishing singletons and geminates is the amount of voicing during the closure. Fig. 6 shows the voicing durations during the various single-

tons and geminates. Note that all segments under investigation are underlyingly voiceless. Voiceless obstruents often have some voicing leaking into their constriction period, and strengthening by gemination should diminish such voicing leaks. As the figure shows, this occurs for the glottal obstruents, but no clear effect is visible for the oral obstruents.

To test this, a linear mixed-effect model was run with voicing duration as the dependent variable and Quantity and Segment as predictors. All predictors were contrast coded (quantity:  $-0.5$  vs  $0.5$ ) and the four-level segment variable was coded as three linearly independent contrasts (Contrast 1: glottal vs oral articulation; Contrast 2: /s/ vs oral stop; Contrast 3: /h/ vs glottal stop). Participants and item were used as random factors with a maximal random effect structure but as uncorrelated random effects. Table 1 shows the outcome of the analysis. There was a main effect of quantity that is qualified by an interaction with place of articulation. The diminishing of the voicing leak under gemination was stronger for the two glottal segments than for the oral segments, which was reflected in a significant regression weight for the interaction term “Glottal versus Oral x Quantity”. When the effect of quantity on voicing leak was tested separately for oral and glottal segments, there was a significant effect for glottal segments (19 ms less voicing during the geminates than during the singleton,  $t = -2.38$ ,  $p = .024$ , with an intercept, the overall mean of voicing duration, at 47 ms) but not for the oral segments (3 ms less voicing duration geminated than during singletons,  $t = -1.06$ ,  $p = 0.30$ , with an intercept at 26 ms).

At this juncture, it might be argued that the difference in alignment of /h/ for singletons and geminates might be an

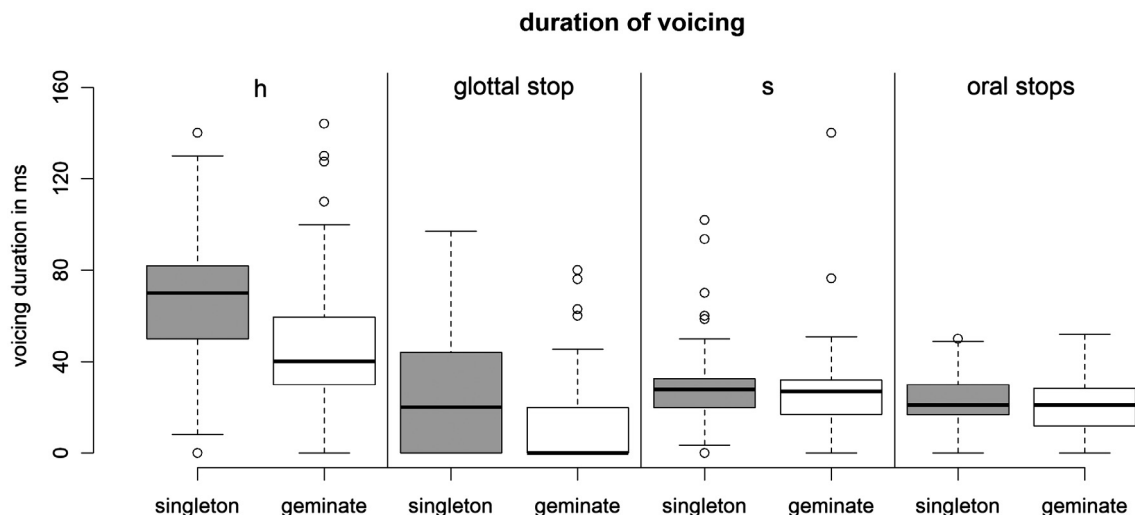


Fig. 6. Boxplots of voicing duration during the construction of the various segments.

artefact of the alignment algorithm. Maybe short and voiced signals get aligned as [h] while long voiceless signals get aligned as [x]. However, we inspected the distribution of voicing and duration within the /h/ geminates. This revealed that there is considerable overlap in duration and voicing between tokens that are transcribed as [h] and [x], which indicates that long voiceless segments can be aligned as [h]. Appendix C (see Fig. A1) shows such an example and contrasts it with an example that was aligned as [x]. This comparison shows that there are spectral differences between these two fricatives. Moreover, it was tested whether alignment as [x] was influenced by duration. We therefore shortened the fricative from a random sample of 25 utterances transcribed with [x] to 60% of its initial duration and performed the alignment again. Invariably, the forced-alignment algorithm still preferred [x] over [h], showing that the differences in how singletons and geminates are aligned are not due to duration differences.

#### 2.2.4. Secondary cues: Vowel duration

Previous work has often focussed on the duration of the surrounding vowels for singleton-geminate distinctions (Kingston et al., 2009). Therefore, we also tested whether the duration of the surrounding vowels depends on quantity. Means per condition and an analysis of the surrounding vowels' duration are presented in the Appendix B (see Tables A1–A4) and show that vowel duration seemed to be independent of quantity of the medial consonant in our CV(C)CVC words. This partly contrasts with the results from Galea (2016), who found an effect of quantity on the preceding vowel but not the following vowel in Maltese. While this requires further investigation, it might be that the vowel shortening also differs by manner as Galea investigated liquids and nasals, and it is possible that vowel shortening is more prominent with these consonants. The shortening observed by Galea was quite subtle (13 ms) and it is unclear whether it was significant. Moreover, it may also be the case that the “vowel melody”<sup>2</sup> in verbs of Semitic origin is a special case in which there might be less phonetic variation.

<sup>2</sup> The vowel melody is a term used for the vowels that a given verb uses to generate the words out of roots. For instance, the root r-q-d (Engl., ‘sleep’) uses the vowel melody /a/-/a/, leading to the form *raqad* (Engl., ‘he slept’) while the root k-t-b (Engl., ‘write’) uses /i/-/e/ to form *kiteb* (Engl., ‘he wrote’).

#### 2.3. Discussion

The results indicate that the differences between singletons and geminates are not uniform between oral and glottal obstruents. For glottal segments, there are clear differences in how singletons and geminates are produced, which are not mirrored in oral segments. This is evident, first, in their different gestures (/h/ vs /x/ and glottal stop vs glottalization) and, second, in the amount of voicing leak into the underlyingly voiceless constrictions.

This is most obvious for /h/ in comparison to /s/. For /h/, the singleton has an average duration of 80 ms, and the mean voicing duration is 68 ms. More than 80% of the singleton /h/ s are fully voiced while the remaining tokens show a relatively uniform distribution of voicing leaks from 10% to 80%. This is in stark contrast with the geminate /h:/, for which only 7% of the tokens are fully voiced. Moreover, there is a clear difference in terms of place of articulation for /h/ but not for /s/. It is not difficult to find a reason for this; /s/ contrasts with both /f/ and /ʃ/, while the closest fricative to /h/ in Maltese is the postalveolar /ʃ/. Speakers can hence easily make use of allophonic variation for /h/ but not for /s/ without endangering phonemic contrasts.

Secondary cues also more strongly distinguish singleton and geminate glottal stops than singleton and geminate oral stops. Singleton glottal stops are mostly realized as glottalization rather than stops while no such differences occur for the oral stops, which reliably surfaced as stops and never as lenited fricatives or flaps. Similarly, we see a clear difference in how much regular voicing there is during underlying glottal stops depending on their quantity. In line with the idea that their voicelessness is strengthened, there is less voicing during the geminates. No such difference, however, is observed for the oral stops.

These data indicate that, even if geminates represent a strengthened articulation compared to a singleton, this strengthening affects different segments differently. Consequently, it is difficult to see how listeners could usefully extract a context-independent “strengthening” feature from the input to gain lexical access. Similarly, if we assume a mora-based account of gemination, listeners face the task of “finding” an extra mora based on disparate perceptual cues for different



segments. This problem is not new at all; Klatt (1989) already noted that the primary problem of feature-based models of word recognition is that features do not have reliable acoustic correlates. How can listeners recognize the feature [TENSE] if the acoustic cues that carry it vary by consonant? Before this becomes a real issue, however, it needs to be shown that these secondary cues do indeed matter. Due to the high collinearity of the different aspects of gemination, it is difficult to show that using the secondary cues may improve classification accuracy. Therefore, Experiment 2 varies secondary cues and duration orthogonally and tests directly how the two factors influence categorization of tokens as singleton and geminates.

### 3. Experiment 2

The purpose of this experiment is to test to what extent the secondary cues found in Experiment 1 affect perception. We therefore tested the classification of seven different duration continua as singletons versus geminates. Two of these continua each used the minimal pairs *wasal-wassal* (Engl., ‘to arrive’ – ‘to give a lift’) and *daħak – daħħak* (Engl., ‘to laugh’ – ‘to make laugh’). We varied the properties of the frication as found in the production study. For /s/, this meant that the centre of gravity was different between the two continua. For /h/, we used a fully voiced /h/ for one continuum and an unvoiced /x/ for the other. The question is whether these secondary cues influence categorization notably.

Three other continua varied the secondary cues for glottal stop using the minimal pair *waqaf-waqqaf* (Engl., ‘to stop’-‘causing to stop’). One continuum contained a full glottal stop and one continuum contained a glottalization-like signal with continuing voicing. A third continuum was intermediate and based on the glottalization-like continuum (see Method for details).

The critical question is to what extent the secondary cues influence categorization of the stimuli as a singleton or geminate. If indeed the strength of secondary cues varies over consonants, we should find stronger differences between the continua based on glottal stop and /h/ than the continua based on /s/.

#### 3.1. Method

##### 3.1.1. Participants

Sixteen native speakers of Maltese participated in this experiment. They were students at the University of Malta aged 18–24, and twelve of them were female. All participants gave written informed consent and were paid for their participation.

##### 3.1.2. Materials

Stimuli were based on the three minimal pairs *wasal-wassal* (Engl., *to arrive-to give a lift*), *daħak – daħħak* (Engl., *to laugh – to cause laughing*) and *waqaf-waqqaf* (Engl., *to stop-causing to stop*). Stimuli were generated via the diphone speech synthesizer MBROLA instead of manipulating the duration of natural utterances to minimize the existence of other secondary cues. Yoshida et al. (2015) had found that such cues exist in natural utterances without being able to pinpoint them. Using

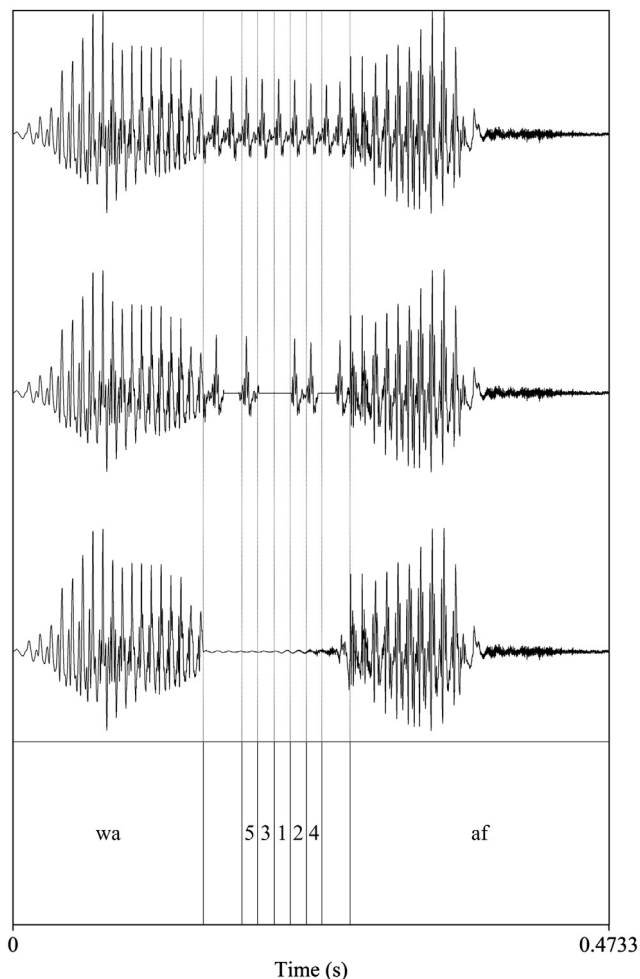
synthesized speech is therefore a way to avoid such possible confounds. For each of the stimuli used, the average duration of each segment was calculated from the recordings generated in Experiment 1. These average segment durations served as synthesis parameters. The stimuli were generated using the diphone synthesizer MBROLA and the voice de6 (i.e., the diphone database used for synthesis) for *wasal-wassal* and *waqaf-waqqaf* and de7 for *daħak – daħħak*. Various voices were presented to a Maltese informant and the stimuli based on these voices did not sound foreign-accented to the informant (note that the prosody of the stimuli was based on the Maltese recordings). The use of a diphone synthesizer allows us to present stimuli that vary in duration but are equated in terms of any unaccounted cues for the singleton-geminate distinction.

Stimuli were generated with the average duration estimated for the geminate and a duration continuum was generated by cutting out parts of these sounds (see Figs. 7 and 8). For the glottal stop, two base stimuli were generated, one with a glottal stop specified (which led to a full stop in the MBROLA output) and one with an /a/ replacing the glottal stop. To mimic glottalization on this vowel, the pitch dropped from 125 Hz to about 80 Hz. That is, the pitch was not held at exactly 80 Hz but varied somewhat to mimic the typically semi-regular creaky voice in case of glottalization. Additionally, the amplitude of these glottal cycles was reduced to 60% of its original, in line with what is typically observed in glottalized periods. At the end of this vowel, the pitch moved back to 110 Hz for the final vowel.

To generate base stimuli with the duration of the original geminate of 113 ms (i.e., the mean duration of the geminate glottal stop in all recordings of the word *waqqaf* in Experiment 1), the middle consonant (and its vowel replacement) were initially generated with a duration of 150 ms and then shortened. Since the splicing requires full glottal cycles to be used, we extracted 116 ms out of the glottalization mimicking stimulus (the value closest to 113 ms that contained complete glottal cycles) and the same duration out of the glottal stop stimulus. The stimulus with glottalization contained ten glottal cycles. For the intermediate base stimulus intermittent glottal cycles, cycles two, four, five, and eight of the stimulus with glottalization were set to zero (see Fig. 7). This gives rise to three stimuli with the typical duration of a geminate.

For the generation of duration continua, these three base stimuli that had a duration of 116 ms were shortened in five steps to 53 ms by cutting out glottal cycles from the middle. Fig. 7 shows the three base stimuli and the different parts consecutively cut out to shorten the duration. The resulting continua hence had medial consonants with a duration that ranged from 53 to 116 ms with steps of 12–14 ms (depending on the duration of the glottal cycles).

The generation of continua for the minimal pair *daħak – daħħak* was also based on cutting out glottal pulses. Again, from a stimulus synthesized with the phone [h], the initial [da] and final [ak] were used. For the middle part, a fully-voiced /h/ with a duration of 135 ms was generated and cut down by two glottal cycles each in five steps of approximately 12 ms. As a consequence, the [h] in the shortest stimulus had a duration of 73 ms, which fits closely with the observed average /h/ durations of 75 ms for the singleton and 138 ms for the geminate of this minimal pair. The same steps were used for a stim-



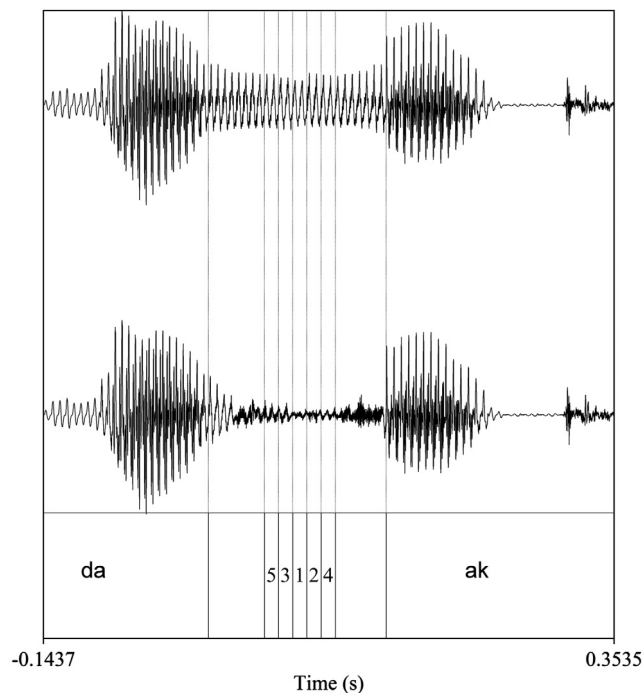
**Fig. 7.** Base stimuli for the glottal-stop continua used in Experiment 2, based on the word *waq(q)af* (Engl., 'to stop' vs 'to make something stop'). The upper waveform shows the glottalization continuum, the middle waveform the intermittent glottalization continuum, and the lower waveform the glottal stop continuum. The numbers indicate the parts cut out (and the order in which they were cut out). The second longest stimuli were generated by cutting out the part labelled "1", the third longest stimuli by cutting out the parts labelled "1" and "2", and so on. [multimedia files available in the [Supplementary data](#) for this article].

ulus with an unvoiced [x]. Fig. 8 shows the base stimuli and the parts cut out to generate a duration continuum.

For the pair *wasal-wassal*, the duration continua did not have to take into account glottal cycles, since the [s] was also phonetically unvoiced. A stimulus was generated with an /s/ of 158 ms, the average duration of the [s] in all recordings of *wasal*. A second /s/ was generated with a centre of gravity that was 400 Hz (the overall mean difference found in Experiment 1) higher using the function *change gender* in Praat (Boersma, 2001). Both stimuli were shortened in five steps of 14 ms, leading to a continuum ranging from 158 to 88 ms, matching the observed mean /s/ durations in the production study for this minimal pair.

### 3.1.3. Procedure

The stimuli were presented to participants in a forced-choice identification task. Each of the 42 stimuli (seven continua with six steps) was presented ten times to each



**Fig. 8.** Base stimuli for the /h/ continua used in Experiment 2, based on the word *dah(h)ak* (Engl., 'to laugh' vs 'to make somebody laugh'). The upper waveform shows the [h] continuum and the lower waveform the [x] continuum. The numbers indicate the parts cut out (and the order in which they were cut out). The second longest stimuli were generated by cutting out the parts labelled "1", the third longest stimuli by cutting out the parts labelled "1" and "2", and so on.

participant. Continua were presented fully intermixed so that participants were presented with a randomized order of the 42 stimuli ten times in a row. After every 50 trials, participants had the opportunity to have a short break. Break trials also informed participants of the number of trials already completed and the number of trials remaining.

On a given trial, the two answer alternatives were presented on the left and right of the screen as written words (with the singleton always on the left) and after 700 ms, the auditory stimulus was played. When participants pressed either the left or right arrow key, the display highlighted the choice made for 500 ms and the next trial was initiated after 800 ms. Experimental sessions lasted about 20–25 min.

### 3.2. Results

Fig. 9 shows the proportion of geminate responses for all seven continua, with different panels for each underlying segment. The upper left panel shows the data for the three continua based on the glottal-stop minimal pair. The results show a clear distinction between the three continua with different segment qualities, and the more stop-like the stimulus, the higher the proportion of geminate responses. This is confirmed by a statistical analysis with a linear mixed-effect model with a binomial linking function with Duration and Segment Quality as fixed factors and participant as a random effect (with a maximal random effect structure, but with no correlations between random effects). Duration was coded so that it was centred around zero, ranging from [−2.5,2.5] in steps of 1. Segment Quality was dummy coded with the partial stop mapped on

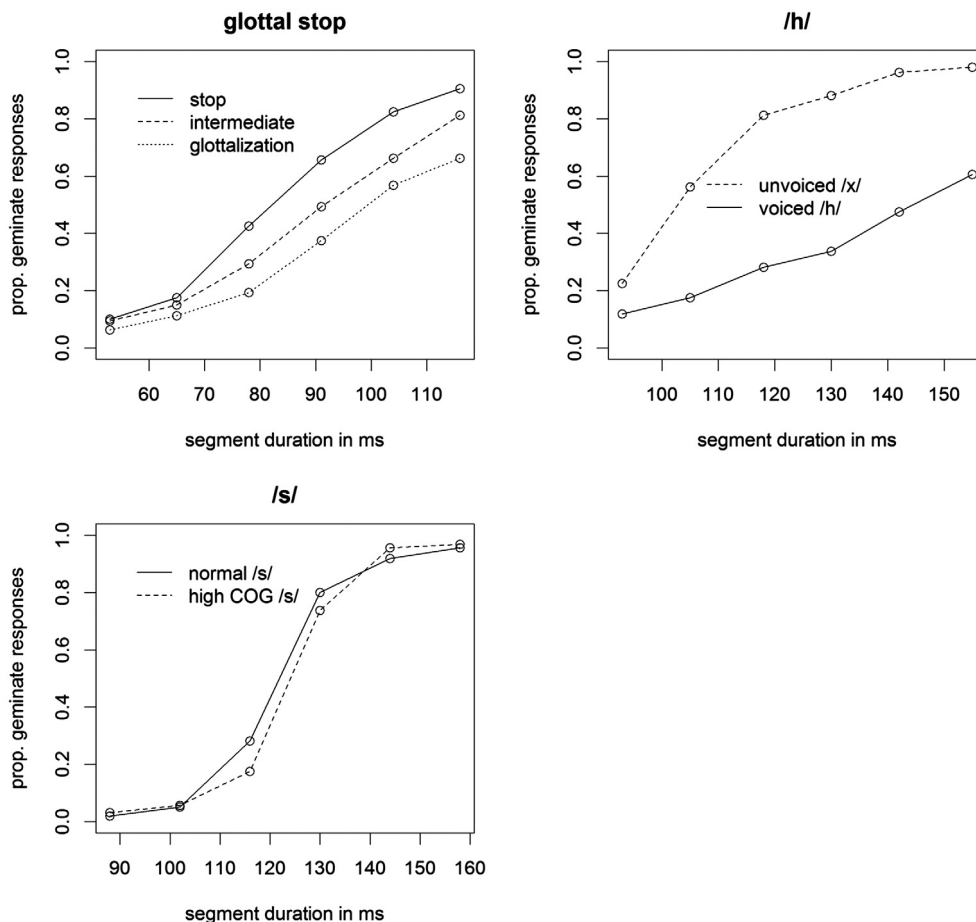


Fig. 9. Proportion of geminate responses for the seven continua organized in different panels for each underlying consonant.

the intercept and dummy variables for the Glottalized and the Full-Stop continuum.

The results (Table 2) show that there was a difference in geminate responses due to the secondary cue that was step-wise. The intermediate condition with a partial stop received more geminate responses than the glottalization continuum but fewer than the full-stop continuum. Additionally, there was a stronger effect of Duration for the full glottal stop continuum than for the other two conditions. This indicates that a short full glottal stop was more likely to be accepted as a singleton than a long glottalization was to be accepted as a geminated glottal stop.

For the analysis of the data from the /h/ continua (upper right panel of Fig. 9), Segment Quality was contrast coded with [h] mapped on -0.5 and [x] mapped on 0.5, so that a positive regression weight would indicate that the “strong” [x] leads to more geminate responses. The intercept, which represents the grand mean of geminate responses in logOdds when contrast coding is used, was at 0.61. There was an effect of Segment Quality, with more geminate responses for the stimuli using [x] ( $b = 3.247$ ,  $SE(B) = 0.429$ ,  $z = 7.573$ ,  $p < .001$ ). Additionally, there was an effect of Duration ( $b = 1.026$ ,  $SE(B) = 0.128$ ,  $z = 8.040$ ,  $p < .001$ ) and an interaction of Segment Quality and Duration ( $b = 0.808$ ,  $SE(B) = 0.140$ ,  $z = 5.738$ ,  $p < .001$ ). The interaction indicates that there is a significant difference in the steepness of the identification functions; the effect of Duration was 0.622 logit units per step for the weak [h]

stimulus but 1.43 logit units per step for the strong [x] stimulus. As the figure shows, this was due to how stimuli with conflicting cues between Segment Quality and Duration were perceived. The short version of the “strong” segment was pre-dominantly accepted as a singleton, but the long version of the weak segment, the voiced [h], was not predominantly accepted as a geminate. Note that this is like the pattern observed with the glottal stop. Consequently, the effect of Duration was larger for the strong version of /h/, with geminate identification rates ranging from 22% to 98%, while the range for the weak version of /h/ was from 12% to 61%.

For /s/, in contrast, a linear mixed-effect model predicting perceived segment quantity based on Duration and Segment Quality revealed only an effect of Duration ( $b = 2.27$ ,  $SE(B) = 0.224$ ,  $z = 10.163$ ,  $p < 0.001$ ) and no effect of Segment Quality ( $B = -0.207$ ,  $SE(B) = 0.157$ ,  $z = -1.318$ ,  $p = 0.188$ ) and no interaction ( $B = 0.134$ ,  $SE(B) = 0.148$ ,  $z = 0.905$ ,  $p = 0.365$ ). As above, the fixed factors were contrast coded so that the observed intercept at -0.03 reflects the overall mean of geminate responses in logOdds.

Given that the effect of Duration was numerically largest for /s/, we tested whether the steepness of the duration continua was significantly larger for /s/ than for the glottal consonants. This analysis only took into account Duration (centred around zero, ranging from [-2.5,2.5]) and Underlying Segment, contrast coded with two contrasts: /s/ versus glottal consonants, and /h/ versus glottal stop. The results revealed an intercept

**Table 2**

Results of the statistical analysis for the categorization data for the three glottal stop continua. Note that the regression weights reflect effects in logOdds space.

Term	B	SE(B)	z	P
(Intercept)	−0.61	0.34	−1.78	0.074
Glottalization	−0.60	0.15	−3.91	<0.001
Glottal stop	0.78	0.30	2.60	0.009
Duration	1.08	0.14	7.91	<0.001
Glottalization: Duration	−0.13	0.09	−1.48	0.140
Glottal stop: Duration	0.35	0.14	2.47	0.014

at  $-0.11$  logOdds and an overall effect of Duration of 1.28 (SE 0.11,  $z = 11.50$ ,  $p < 0.001$ ). Given the use of contrast coding these values reflect the overall mean of geminate responses in logOdds and the mean effect of duration over all three continua. The analysis also showed that both contrasts led to significant interactions with Duration, with a stronger effect of Duration for /s/ than for the two glottal consonants ( $B = -0.906$ ,  $SE(B) = 0.115$ ,  $z = -7.874$ ,  $p < 0.001$ ) and a smaller but still significant difference between /h/ and glottal stop ( $B = -0.358$ ,  $SE(B) = 0.065$ ,  $z = -5.516$ ,  $p < 0.001$ ), with a smaller effect of Duration for /h/ than for the glottal stop. This indicates that the smaller the effect of secondary cues (/s/ < /ʔ/ < /h/), the larger the effect of duration (/s/ > /ʔ/ > /h/).

### 3.3. Discussion

The data clearly indicate that there are strong influences on quantity perception that are triggered by cues other than the duration of the consonants to be geminated. In some cases, there are no clear perceptual switches based on duration alone, even though our endpoints are modelled on typical durations for singleton and geminates based on the production data. For glottal stop and /h/, a signal with the typical duration of a geminate but secondary cues for a singleton only receives 60% and 66% geminate responses, respectively. This contrasts with previous findings that secondary cues are not used at typical durations for singleton and geminates (Hankamer & Lahiri, 1988; Hankamer, Lahiri, & Koreman, 1989, for evidence from Turkish and Bengali). Those studies found that singletons with a manipulated duration that was typical for a geminate are convincing geminates, which is why, as these papers argued, secondary cues can easily be disregarded. This does not seem to be the case for Maltese glottal geminates.

Interestingly, our data also indicate how these apparently conflicting data may come about. The strength of secondary cues varies strongly over target consonants. Secondary cues are strongest for /h/, of medium strength for glottal stop, and absent or weak for /s/. The use of secondary cues may be related to how much variation there is in singletons<sup>3</sup>. As already noted, there is an allophonic variation in which /h/ tends to be produced either as a glottal fricative or with an additional oral constriction, especially in positions in which it is not perceptually salient. Indeed, one of the typical demonstrations in introductory phonetic classes is that playing the word “Hanna” backwards gives rise to the percept of “Anna”, since most of the acoustic cues for /h/ in such a reversed utterance are not perceptually salient. Therefore, a post-vocalic /h/ is enhanced in Maltese with an additional oral constriction. This enhanced

variant is then also used for the geminate in intervocalic position. A similar situation may arise, for instance, when /r/ is geminated in Italian. The default allophones for singleton /r/ in Italian are a trill for the onset position and a tap for the intervocalic position (Ladefoged & Maddieson, 1996, pp. 220–221). If the /r/ is geminated in intervocalic position, the “stronger” allophone, the trill usually not used in this position comes into play.<sup>4</sup> This is analogous to the current data with Maltese /h/, in which the allophone mostly used for the coda position is used in intervocalic position in case of a geminate. The strength of secondary cues for a given geminate may hence be related to how much variation there is for the singleton over different phonetic environments.

The data also indicate that the strength of secondary cues leads to differential effects for duration. For /s/, we find a very clear and steep identification function based on duration; for glottal stop, we find shallower slopes; and for /h/ the identification functions are more linear than s-shaped. This latter finding also indicates that it is unlikely that there are strong secondary cues for /s/ that were not measured. This would certainly be a possibility; this project did not aim to find all potential secondary cues, but rather focused on the idea that some secondary cues may be strong for some consonants and weak for others. While it is possible that some cues may be found in other measures, the steep sigmoid identification functions found for /s/ indicate that it is unlikely that there are other cues which have a lot to contribute to quantity perception.

Finally, the current data indicate that the secondary cues affect quantity perception in an asymmetric way. For both glottal stop and /h/, secondary cues affected stimuli at the short and long end of the duration continua differently. The categorization of short consonants as singletons was not influenced strongly by the secondary cues, but long consonants were strongly influenced by secondary cues. A long consonant was then only consistently accepted as a geminate if it had a long duration and appropriate secondary cues. This asymmetry would be in line with the idea proposed by Kotzor et al. (2016) that geminates need more perceptual evidence than singletons because they are attached to an additional mora, which must be supported by bottom-up input. They based this on their finding that words in which a geminate is mispronounced as a singleton (as an experimental manipulation) do not achieve lexical access, while the opposite mispronunciation (a singleton mispronounced as a geminate) does. However, there is an alternative to this representation-based assumption to consider. Bonte, Mitterer, Zellagui, Poelmans, and Blomert (2005) had shown that frequency of usage may often explain effects apparently caused by asymmetric representation. Therefore, we estimated the frequency of singleton and geminates for /h/ and glottal stop using the Corpus Malti (v2.0), which is a collection of Maltese texts with over 130 million words. Though the corpus is based on written language, it does allow an estimate of how often various sounds are used as Maltese orthography is relatively shallow. In fact, the sounds /h/, /h:/, /ʔ/ and /ʔ:/ in Maltese are always written as ħ, ħħ, q, and qq, respectively. The corpus indicated that the singleton glottal stop is about 14 times (i.e., 1400%) as frequent

<sup>4</sup> For this reason, the /r/ is sometimes omitted from studies of gemination in Italian (Payne, 2005). Importantly, such design choices then reinforce the assumption that gemination is a relatively uniform process for all segments in a given language.

<sup>3</sup> This line of reasoning was suggested by an anonymous reviewer.

as the geminate glottal stop (frequency per million of 24,584 for singleton glottal stop but only 1760 occurrences of geminate glottal stop per million). An even stronger bias for the singleton is found for /h/, with the singleton being more than 25 times (i.e., 2500%) as frequent (92,307 singletons and 3578 geminates per million words). The asymmetry found in the data may hence be simply the consequence of an optimal classification algorithm that follows the credo that “extraordinary claims require extraordinary evidence”. Because glottal geminates are relatively rare to begin with, an optimal classification algorithm will simply require good evidence for its occurrence. Consequently, for /h/ and glottal stop, a stimulus is only considered a good example of a geminate if it has an adequate duration *and* contains adequate secondary cues.

#### 4. General discussion

This paper aimed at investigating whether geminates can have “unique” secondary cues. “Unique” here means that these cues cannot be easily understood as some form of strengthening of the articulatory gestures of the singleton and are moreover specific to the geminated consonant in question.

Experiment 1 showed that this was the case in production. For glottal stops, the data confirmed the informal observation by Ladefoged and Maddieson (1996) that singleton glottal stops are usually not produced as stops but rather as a glottalization; stops are only regularly observed for geminates. In this, the glottal stops differed from oral stops, which were produced as stops in both cases (i.e., singleton and geminate). The glottal fricative /h/ also showed gestural variation triggered by gemination. As a singleton, it was mostly produced as a phonetically fully voiced fricative. As a geminate, it was produced as a voiceless fricative—with some voicing leak from the previous vowel—and with an additional oral constriction. The oral fricative /s/ showed no such pattern; the amount of voicing leak into the underlyingly voiceless segment was equivalent for singletons and geminates, and only a small acoustic difference was found in the centre of gravity during friction.

Experiment 2 hence tested the importance of the strong secondary cues for glottal consonants and the weak cues for /s/ in perception. The results showed that the importance of secondary cues on quantity perception differs between segments. For /s/, duration is the main and by far most important cue. For the glottal stop, duration plays a smaller but still quite dominant role, while for /h/ a convincing geminate percept cannot be supported by a geminate-like duration alone. Listeners also require a secondary constriction to perceive a long consonant as a convincing geminate.

These results challenge current concepts of how singletons and geminates are represented in speech processing, whether as a prosodic weight difference (Kotzor et al., 2016) or due to an additional tensification (Ridouane, 2010) or with a feature [LONG]. These approaches predict that the acoustic and articulatory differences between singletons and geminates are relatively uniform across consonants. The current data provide an existence proof that this might not always be the case. Instead, the current data suggest that singleton and geminates, for the purpose of perception and production of the contrast, may be segments in their own right that cannot be

decomposed into features. While it may be argued that, on an abstract phonological level, there is still something common to all geminates, at the phonetic implementation stage, and consequently for the listeners, geminate consonants seem to have their own “private life” and have properties that are not directly predictable from the singleton. Given the current data, listeners face an easier task in recognizing singleton and geminate glottal consonants with independent prelexical units, rather than recognizing the same consonant and an additional timing unit or duration feature. This strongly reduces that amount of variance the listener must deal with, but at the expense that a wider range of prelexical representations are required. However, recent research has suggested that listeners have more units for prelexical abstraction than there are phonemes in their respective languages (Mitterer, Scharenborg, & McQueen, 2013, see below for details).

One possible way to view this is to compare consonant quantity with vowel quantity, which may not be that different from each other. Short and long vowels often have unreducible properties and each language makes its own choices how to deal with each segment. An instructive case is the one of the closely related languages German and Dutch, both of which have long/tense and short/lax vowels, but with their specific phonetic properties. Dutch high long vowels tend to be diphthongized, but for the low vowel /a/ there is a strong spectral difference between tense and lax vowels, and no diphthongization. German has a structurally similar vowel system, but has no or little diphthongization of long vowels, and the two low vowels differ only in duration and not in quality. Each segment may hence acquire its own secondary characteristics, and as a consequence make abstract features not very functional for speech processing. In other words, gemination also involves language-specific and even segment-specific phonetic knowledge (Kingston & Diehl, 1994).

As always, it is possible to save any theoretical framework by making ad-hoc assumptions (see Lakatos, 1968 and the tale of the stubborn Newtonian). To accommodate the current data under the assumption that lexical access is based on features, the following ad-hoc assumptions would be possible: all geminates make use of the same features as the singleton plus the feature [LONG] or [TENSE]. In addition, some geminates make use of additional and unpredictable feature modifications. The geminate version of /h/ could then take on an additional constriction feature. The geminate version of the glottal stop could have a different specification of a continuant feature, which usually distinguishes stops from fricatives. The singleton glottal stop might be specified as [+CONT] but the geminate might be specified as [−CONT]. The problem with this form of ad-hoc assumption is twofold. First of all, it undermines the *raison d'être* for featural accounts, which is to explain a wide range of segments as a combination from a small set of features that combine independently. Secondly, and more gravely, it immunises the featural approach against any potential problematic evidence, because whenever a segment has properties that cannot be explained from the combination of features, one can simply think of which additional features this particular segment may have.

It nevertheless remains possible that, at a phonological planning level, gemination occurs through an additional timing unit based on a feature such as [LONG] or [TENSE]. However,

at the phonetic implementation level, additional cues are used depending on the segment. Moreover, the current data do not speak to the formal phonological representation of the singleton-geminate contrast. Formal phonological theory has mostly focussed on explaining phonotactic patterns in the languages of the world, but it is not straightforward to link this to phonetic data, an example of the Quine-Dunham problem. Therefore, the current paper focussed on theories that are explicit in how to make these linking assumptions. Kotzor et al. (2016), for instance, assume that geminates have an extra mora which is expressed through a longer duration. Accounts based on Articulatory Phonology (Ridouane, 2010) assume that geminates are more forceful versions of their singleton counterparts and hence are not only longer, but uniformly stronger. The current paper clearly challenges these accounts, which assume that the distinction should be expressed relatively uniformly across the board.

Muller (2001) had already provided some tentative evidence that geminates may be expressed differently across different languages, with VOT playing a more important role in Cypriot Greek than in other languages. However, differences between languages may still be explained within a uniform framework. For instance, Cho and colleagues (Cho & McQueen, 2005; Cho et al., 2007) showed that the acoustic consequences of prosodic strengthening can differ across languages, yet still be governed by the same principles. VOT in voiceless stops may increase or decrease with prosodic strengthening, depending on how exactly a voiceless stop is specified in each language. Similarly, differences in secondary cues for gemination between languages may be explained as differences in the phonetic specification of the respective singleton segments, which get enlarged by gemination. In this context, two aspects of the current data are important: First, the data show that some strengthening processes can differ by segment within a language, and these strengthening processes cannot be easily explained as a strengthening of the singleton. Secondly, not all segments in a given language have strong secondary cues. The fact that for /s/, duration seems to be the all-deciding cue but not for /h/ and /ʔ/ indicates that listeners cannot recognize geminates based on some shared features, but are forced to use different representations for each segment.

This does not preclude that more abstract representations have no cognitive reality. Indeed, assuming an abstract feature for long vowels in Dutch and German does make it easier to account for the permissible syllable structures in Dutch and German (syllables containing lax/short vowels must have a coda). Moreover, on a morphological level, gemination for Maltese verbs has the common characteristic of leading to a causative meaning, independent of underlying consonant. However, at a phonetic level in perception and production, segments as combinations of features take on irreducible properties, and, as a consequence, at a perceptual level, such features are not very useful for decoding the message.

This gives rise to the question how to account for data by Kotzor et al. (2016) that seem to support a unitary, across-the-board relation between singleton and geminates. Kotzor et al. argued that their priming data from Bengali singletons and geminates support a moraic account. They found that geminate substitutions of underlying singletons (i.e., producing

intended *waqaf* as *waqqaf*) primed both the singleton and the geminate, while geminate substitutions (i.e., producing intended *waqqaf* as *waqaf*) only primed the singleton but did not lead to priming for the geminate. They argued that this favours a moraic account because the geminate contains all information for the singleton, but the singleton misses the extra moraic weight necessary to support the perception of a geminate. This finding may, however, also be explained by the assumption of optimal categorization (Mitterer, 2011). In the discussion of Experiment 2, it was already mentioned that, in Maltese, singletons are more frequent than geminates. Given that geminates are often viewed as *marked*, and markedness is often related to (low) frequency, a higher frequency of singletons when compared with geminates might in fact be a general tendency. Interestingly, in a corpus analysis of Japanese, Tajima (2013) found that, despite clearly different mean durations between singletons and geminates, at all durations, a singleton is more likely than a geminate. Due to the higher base rate of singletons and a larger standard deviation of their duration distribution, even quite long consonants are more likely to stem from underlying singletons than geminates. Optimal categorization of such distribution predicts exactly the data as reported by Kotzor et al. (2016): At short durations, the perceiver can be relatively sure that the underlying segment is a singleton, but at long durations, both are possible. Obviously, this would depend on the exact distribution of singleton and geminate duration in a given language (one cannot assume that the findings for Japanese would replicate in Bengali, see Evans & Levinson, 2009), and this may also explain why the results vary over languages: Tagliapietra and McQueen (2010) performed similar priming experiments as Kotzor et al. in Italian and found the opposite pattern. Singletons can prime geminates, and to some extent even better than geminates prime singletons. The existence of such differences indicates that a one-size-fits-all explanation for gemination is unlikely to be successful. There are clear differences between languages; languages differ in how gemination is implemented (Kingston et al., 2009) and in priming effects caused by singleton and geminate mispronunciations (Italian: more priming for geminates; Bengali: more priming for singletons). The current data indicate that even within a language, different geminate consonants may differ from one another in how they are implemented.

As the current data question the use of features in speech processing, they dovetail well with another line of research that recently also questions whether featural decomposition is functional for speech processing. A perceptual learning paradigm has shown that listeners make use of abstract units in speech perception (Norris, McQueen, & Cutler, 2003): If listeners repeatedly hear a sound that is ambiguous between /s/ and /f/ at the end of words that canonically end on /s/ (such as *police*), they learn that this ambiguous sound is an /s/ and are able to generalize this to other words (hence, they interpret [nai<sup>s</sup>/i] as *nice* rather than *knife*). This raises the question what type of unit participants learn about in this kind of experiment. Mitterer and colleagues tested this in a series of experiments. First, they showed that abstract phonemes are unlikely targets, as learning does not generalize from one allophone to another (Mitterer et al., 2013; Reinisch, Wozny, Mitterer, & Holt, 2014). If Dutch listeners learn to interpret a sound that is ambiguous

between an approximant /r/ and a dark /l/ as an /r/, they do not also perceive ambiguous sounds between a trilled /r/ and a light /l/ as an /r/. In further studies, it was tested how features fare as a potential unit that listeners learn about. If learning about a sound between /s/ and /f/ is based on features, they should learn about the place feature and generalize to other sounds that also differ in place (e.g., /n/ and /m/). Numerous studies have shown that such generalizations do not occur (Mitterer, Cho, & Kim, 2016; Mitterer & Reinisch, 2017; Reinisch & Mitterer, 2016; Reinisch et al., 2014) unless there is a strong acoustic overlap between exposure and generalization items (Kraljic & Samuel, 2006; Mitterer et al., 2016). Learning therefore seems to be tied to specific allophones (Mitterer et al., 2013) but can also occur for even larger units such as highly-frequent syllables (Poellmann, Bosker, McQueen, & Mitterer, 2014).

To conclude, the current data show that geminates may not always be simply long versions of the singleton consonants, but can take on their own irreducible properties which may not only be language- but also consonant-specific. This makes it unlikely that listeners would always recognize them as longer or prosodically stronger versions of the singleton consonants. Instead, they may be consonants in their own right and be recognized through relatively independent prelexical representations, comparable to long and short vowels that differ in both durational and spectral measures.

#### Author note

This work was supported by a University of Malta Research Grant to the first author. This paper has benefitted from the careful reading and comments by Adam Ussishkin, Maria Tabain, and three anonymous reviewers.

#### Appendix A.

##### Stimulus material in Experiment 1.

**Table A1**

Verb pairs and the respective target sentences in which they were elicited in Experiment 1.

ċahad-ċahhad	M'ilux Matthew ċahad li saraq il-flus. (Recently, Matthew denied that he stole the money)	sehet-sehhet	F'nofs ta lejl is-sahhar sehet lil-principessa. (In the middle of the night, the wizard cursed the princess)
dahak-dahhak	atul ir-Randan il-missier ċahhad lil ibnu mill-helu. (During lent, the father forbade his son from having sweets)	wehel-wahhal	Kmieni filghodu spluzzjoni sehhet fil-qorti. (Early in the morning, there was an explosion at the court)
dahal-dahhal	llum Daniel dahak fir-recta. (Today, Daniel, laughed during a play)		Dalghodu Daniel wehel fil-lift. (This morning, Daniel got stuck in the elevator)
lahaq-lahhaq	lbierah Daniel dahhak lil-udjenza. (Yesterday, Daniel made the audience laugh)	faqqa'-faqqa'	Waranofsinhar Daniel wahhal l-inkwatri. (In the afternoon, Daniel put up the frames)
rohos-rahhās	Tard billejl Daniel dahal fis-sodda. (Late at night, Daniel got into the bed)		Il-lejl li ghadda Matthew faqqa' il-karozza. (Last night, Matthew crashed the car)
sahan-sahhan	Waranofsinhar Matthew dahhal l-ikel fil-forn. (In the afternoon, Matthew put the food into the oven)	laqa'-laqqa'	Matul il-lejl is-silg faqqa' mal-hgieg. (During the night, the ice snapped against the windows)
	Bilkemm Matthew lahaq tal-linja. (Just in time, Matthew reached the bus)	laqat-laqqat	Fl-ahhar Matthew laqa' l-ballun. (At last, Matthew caught the ball)
	Is-sena li ghaddiet l-imghallem lahaq lil ibnu minflok. (Last year, the boss hired his son instead of him)		Is-Sibt li ghadda Matthew laqqa' lit-tfajla ma' ommu. (Last Saturday, Matthew introduced his girlfriend to his mother)
	Jumejn ilu il-petol rahas b'zewg centezmi. (Two days ago, the gasoline went down by two cents)	naqas-naqqas	Filghodu Daniel laqat il-vazun. (In the morning, Daniel knocked off the vase)
	Ix-xahar li ghadda il-gvern rahhas l-ilma. (Last month, the government made the water bills cheaper)		Filghaxija Matthew laqqat l-ikel kollu. (In the evening, Matthew ate up all the food)
	Dalghodu il-mutur sahan fit-tigrija. (This morning the motorbike engine heated up in the race)	naqax-naqqax	llum l-elettriku naqas b'10%. (Today, electricity went down by 10%)
	Filghaxija Matthew sahan l-ikel. (In the evening, Matthew heated the food)		Sena ilu Daniel naqqas ix-xorb. (Last year, Daniel reduced his drinking)
		raqad-raqqad	lbierah, Matthew naqax l-statwa. (Yesterday, Matthew carved a statue)
			Waranofsinhar, Matthew naqqax il-kolonna. (In the afternoon, Matthew carved a pillar)
		waqaf-waqqaf	Il-lejl li ghadda Matthew raqad fuq is-sufan. (Last night, Matthew slept on the sofa)
			Waranofshinar Matthew raqqad lit-tifla. (In the afternoon, Matthew put his daughter to sleep)
		waqqa'-waqqa'	lbierah, Matthew waqaf quddiem il-hanut. (Yesterday, Matthew stopped in front of the store)
			Dalghodu, il-pulizija waqqaf il-karozza. (This morning, the police stopped the car)
		basar-bassar	llum Matthew waqqa' it-taraġ. (Today, Matthew fell down the stairs)
			lbirahtlula Daniel waqqa it-tazza. (Yesterday, Daniel dropped the glass)
		haseb-hasseb	Jumejn ilu Daniel basar il-maltemp. (Two days ago, Daniel predicted the thunderstorm)
			Xahar ilu Matthew bassar in-numri tal-lotterija. (Last month, Matthew predicted the lotto numbers)
		kesah-kessah	Reċentament Matthew haseb dwar Amelia. (Recently, Matthew thought about Amelia)
			llum, Matthew hasseb lil-ghalliema. (Today, Matthew worried his teacher)
		kiser-kisser	Filghaxija Daniel kesah fil-gallarija. (In the evening, Daniel felt cold in the balcony)
			lbierah Il-fan kessah il-kamra. (Yesterday, the fan cooled the room)
		nesa-nessa	Il-lejl li ghadda Daniel kiser saqajh. (Last night, Daniel broke his leg)
			Ftit ilu Matthew kisser it-televixin. (A short while ago, Matthew broke the television)
		qasam-qassam	lbirahtlula Matthew nesa ic- cwieviet. (Yesterday, Matthew forgot his keys)
			Jumejn ilu Matthew nesa l-ktieb lil-Anita. (Two days ago, Matthew made Anita forget her book)
		resaq-ressaq	Filghodu Daniel qasam il-triq. (In the morning, Daniel crossed the street)
			F'nofs ta' lejl Daniel qassam ix-xampanja. (In the middle of the night, Daniel distributed the champagne)
		wasal-wassal	Mal-ewwel Matthew resaq fil-kantuniera. (Immediately, Matthew moved into the corner)
			Waranofsinahr Matthew ressaq il-karozza. (In the afternoon, Matthew moved the car)
		rikeb-rikkeb	Filghaxija Daniel wasal id-dar. (In the evening, Daniel arrived home)
			lbierah Daniel wassal lill-Anna. (Yesterday, Daniel gave a lift to Anna)
		rika-rikka	lbirahtlula Matthew rikeb zieme! (Yesterday, Matthew rode a horse)
			Ftit ilu Matthew rikkeb lit-tifla fuq il-mutur. (A while ago, Matthew let his daughter ride on the motorbike [with him])
		beqa-bekka	lbierah Daniel beka fil-funeral. (Yesterday, Daniel cried during the funeral)
			Waranofsinhar Daniel beka lit-tifel. (Yesterday, Daniel made his son cry)

fetah-fettah	Ftit ilu Matthew fetah il-bieb. (A short while ago, Matthew opened the door) Illum Matthew fetah il-glekk. (Today, Matthew enlarged his jacket)
kotor-kattar	Mal-ewwel ir-rummen kotor fil-harifa. (Immediately, the pomegranates spread in the autumn) Ix-xahar li ghadha Matthew kattar il-patata. (Last month, Matthew grew the potatoes)
nitef-nittef	Waranofshinar Daniel nitef it-tigieġa. (In the afternoon, Daniel plucked the chicken) F'daqqa wahda Matthew nitef il-fjura. (Suddenly, Matthew plucked the flower)
niten-nitten	Jumejn ilu Il-hut niten fil-platt. (Two days ago, the fish started smelling on the plate) Filghaxija iż-żibel nitten l-uffiċċju. (In the evening, the garbage caused the office to smell)
qata-qatta'	Jumejn ilu Daniel qata' l-hobz. (Two days ago, Daniel cut the bread) Ilbierah Matthew qatta il-ktieb. (Yesterday, Matthew ripped the book to shreds)
fetaq-fettaq	Il-lejl li ghadha Matthew fetaq il-qmis. (Last night, Matthew ripped the shirt) Mal-ewwel Matthew fetaq dwar id-dokumenti. (Immediately, Matthew fussed about the documents)

**Appendix B.**

**Effects of gemination on vowel duration**

**Table A2**

Table of durations of the vowels preceding and following the singleton/geminate (in ms).

Consonant	Preceding vowel		Following vowel	
	Singleton	Geminate	Singleton	Geminate
/h/	63	60	88	84
/s/	76	77	89	86
Oral stop	69	75	106	105
Glottal stop	67	81	94	94

**Table A3**

Results of a linear mixed-effect model with preceding vowel duration as dependent variable and Consonant (contrast coded into: Glottal vs Oral consonant, /s/ vs oral stop, /h/ vs glottal stop) and Quantity as independent variable.

Term	Estimate	t	p
(Intercept)	69.317	23.527	<0.001
Quantity	4.12	1.16	0.25
Glottal vs Oral	-7.63	-1.43	0.16
/s/ vs oral stop	-6.41	-0.86	0.39
/h/ vs glottal stop	4.44	0.62	0.54
Quantity x Glottal vs Oral	2.24	0.31	0.76
Quantity x /s/ vs oral stop	3.45	0.33	0.74
Quantity x /h/ vs glottal stop	16.86	1.72	0.09

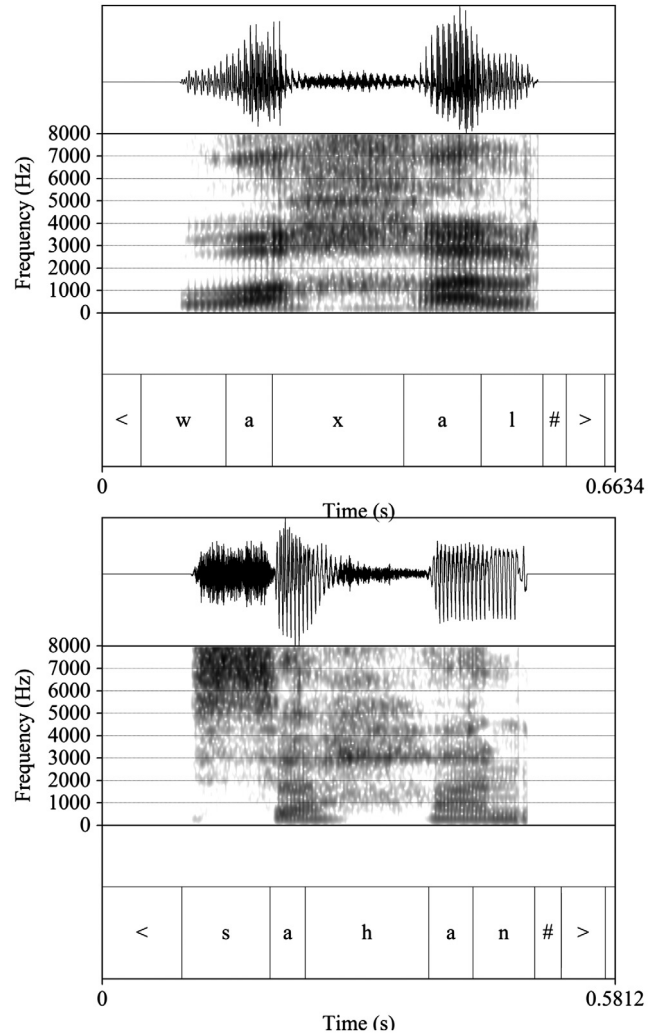
**Table A4**

Results of a linear mixed-effect model with following vowel duration as dependent variable and Consonant (contrast coded into: Glottal vs Oral consonant, /s/ vs oral stop, /h/ vs glottal stop) and Quantity as independent variable.

Term	Estimate	t	p
(Intercept)	94.67	20.98	<0.001
Quantity	-1.67	-0.31	0.76
Glottal vs Oral	-6.46	-0.86	0.39
/s/ vs oral stop	16.04	1.49	0.14
/h/ vs glottal stop	6.34	0.61	0.55
Quantity x Glottal vs Oral	0.16	0.02	0.99
Quantity x /s/ vs oral stop	5.05	0.34	0.74
Quantity x /h/ vs glottal stop	5.05	0.36	0.72

**Appendix C.**

**Forced alignment examples for [h] and [x]**



**Fig. A1.** Examples of long (>150 ms), mostly voiceless signals aligned as /h/ and /x/, respectively. The spectrogram shows that there are clear spectra differences between the two fricatives.

**Appendix D. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.wocn.2017.09.003>.

**References**

Azzopardi-Alexander, M., & Borg, A. (1997). *Maltese* (1st ed.). London, New York: Routledge.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.

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

Bonte, M. L., Mitterer, H., Zelligui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, 116(12), 2765–2774. <https://doi.org/10.1016/j.clinph.2005.08.012>.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33(2), 121–157. <https://doi.org/10.1016/j.wocn.2005.01.001>.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35(2), 210–243.



- Draxler, C., & Jänsch, K. (2004). SpeechRecorder – A universal platform independent multi-channel audio recording software. In: *In Proceedings. of 4th intl. conference on language resources and evaluation*.
- Embick, D., & Poeppel, D. (2015). Towards a computational(ist) neurobiology of language: Correlational, integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, 30(4), 357–366. <https://doi.org/10.1080/23273798.2014.980750>.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429–448. <https://doi.org/10.1017/S0140525X0999094X>.
- Galea, L. (2016). *Syllable structure and gemination in Maltese*. Cologne, Germany: University of Cologne.
- Hamzah, M. H., Fletcher, J., & Hajek, J. (2012). An acoustic analysis of release burst amplitude in the Kelantan Malay singleton/geminate stop contrast. In: *Proceedings of the 14th Australasian international conference on speech science and technology* (pp. 85–88). Retrieved from <[http://www.academia.edu/download/20435164/SST\\_2012\\_KM\\_Academia.pdf](http://www.academia.edu/download/20435164/SST_2012_KM_Academia.pdf)>.
- Hamzah, M. H., Fletcher, J., & Hajek, J. (2016). Closure duration as an acoustic correlate of the word-initial singleton/geminate consonant contrast in Kelantan Malay. *Journal of Phonetics*, 58, 135–151. <https://doi.org/10.1016/j.wocn.2016.08.002>.
- Hankamer, J., & Lahiri, A. (1988). The timing of geminate consonants. *Journal of Phonetics*, 16–327.
- Hankamer, J., Lahiri, A., & Koreman, J. (1989). Perception of consonant length: Voiceless stops in Turkish and Bengali. *Journal of Phonetics*, 17(4), 283–298.
- Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2), 253–306.
- Hume, E., Rose, D., & Spagnol, M. (2014). Maltese word-final singleton-geminate contrasts: An information-theoretic approach. In R. Kager, J. Grijzenhout, & K. Sebregts (Eds.), *Where the principles fail: A Festschrift for Wim Zonnefeld* (pp. OTS–Utrecht). UK: Sage.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3), 419–454. <https://doi.org/10.2307/416481>.
- 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. <https://doi.org/10.1016/j.wocn.2009.03.007>.
- Klatt, D. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.
- 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. <https://doi.org/10.1177/0023830915580189>.
- Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, 13, 262–268. <https://doi.org/10.3758/BF03193841>.
- Ladefoged, P., & Maddieson, I. (1996). *Sounds of the world's languages*. Oxford: Blackwell Publishers.
- Lakatos, I. (1968). Criticism and the methodology of scientific research programmes. *Proceedings of the Aristotelian Society*, 69, 149–186. <https://doi.org/10.2307/4544774>.
- Lubbers, M., & Torreira, F. (2013). *Praatalign: An interactive Praat plug-in for performing phonetic forced alignment*. Retrieved from: <<https://github.com/dopefish/praatalign>>.
- Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 496–513. <https://doi.org/10.1037/a0020989>.
- Mitterer, H., Cho, T., & Kim, S. (2016). What are the letters of speech? Testing the role of phonological specification and phonetic similarity in perceptual learning. *Journal of Phonetics*, 56, 110–123. <https://doi.org/10.1016/j.wocn.2016.03.001>.
- Mitterer, H., & Reinisch, E. (2017). Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops. *Language, Cognition and Neuroscience*, 1–15. <https://doi.org/10.1080/23273798.2017.1286361>.
- Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2), 356–361. <https://doi.org/10.1016/j.cognition.2013.07.011>.
- Muller, J. S. (2001). *The phonology and phonetics of word-initial geminates*. The Ohio State University. Retrieved from: <[https://etd.ohiolink.edu/pg\\_10?0::NO:10:P10\\_ACCESSION\\_NUM:osu1364226371](https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu1364226371)>.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238. [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9).
- Payne, E. M. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association*, 35(2), 153–181. <https://doi.org/10.1017/S0025100305002240>.
- Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/and glottal stop. *Papers in laboratory phonology II: Gesture, segment, prosody*, 90–117.
- Poellmann, K., Bosker, H. R., McQueen, J. M., & Mitterer, H. (2014). Perceptual adaptation to segmental and syllabic reductions in continuous spoken Dutch. *Journal of Phonetics*, 46, 101–127. <https://doi.org/10.1016/j.wocn.2014.06.004>.
- Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics*, 55, 96–108. <https://doi.org/10.1016/j.wocn.2015.12.004>.
- Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, 45, 91–105. <https://doi.org/10.1016/j.wocn.2014.04.002>.
- Ridouane, R. (2010). Geminates at the junction of phonetics and phonology. In C. Fougereon, B. Kuehnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 61–90). Berlin, Boston: De Gruyter. doi: 10.1515/9783110224917.1.61.
- Rietveld, T., & Van Heuven, V. J. (1997). *Algemene fonetiek*. Bussum, The Netherlands: Coutinho.
- Schiel, F. (1999). *Automatic phonetic transcription of non-prompted speech*. Retrieved from: <[https://epub.uni-muenchen.de/13682/1/schiel\\_13682.pdf](https://epub.uni-muenchen.de/13682/1/schiel_13682.pdf)>.
- Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, 39(1), 96–109. <https://doi.org/10.1016/j.wocn.2010.11.006>.
- Strunk, J., Schiel, F., & Seifart, F. (2014). Untrained forced alignment of transcriptions and audio for language documentation corpora using webmaus. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, & J. Mariani, et al. (Eds.), *Proceedings of the ninth international conference on language resources and evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014* (pp. 3940–3947). European Language Resources Association (ELRA).
- 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. <https://doi.org/10.1016/j.jml.2010.05.001>.
- Tajima, K. (2013). *Are phonemic length contrasts exaggerated in Japanese infant-directed speech?* Presented at the ICPP3 (3rd international conference on phonetics and phonology), Tokyo, Japan.
- Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130, 1606–1617.
- Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon Press.
- 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. <https://doi.org/10.1016/j.wocn.2014.12.006>.