Journal of Memory and Language 108 (2019) 104034

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The glottal stop between segmental and suprasegmental processing: The case of Maltese

Holger Mitterer[a], Sahyang Kim[b], Taehong Cho[c,*]

[a] Department of Cognitive Science, University of Malta, Msida, Malta
[b] Department of English Education, Hongik University, Seoul, Republic of Korea
[c] Hanyang Institute for Phonetics and Cognitive Science, Department of English Language and Literature, Hanyang University, Seoul 04763, Republic of Korea

ARTICLE INFO

Keywords:
Prosodic processing
Segmental processing
Glottal stop
Maltese
Eye-tracking and gating
Phonemic vs. epenthetic
Spoken word recognition

ABSTRACT

Many languages mark vowel-initial words with a glottal stop. We show that this occurs in Maltese, even though the glottal stop also occurs as a phoneme in Maltese. As a consequence, words with and without an underlying (phonemic) glottal stop (e.g., a glottal stop-zero minimal pair *qal* /ʔɑ:l/ vs. *ghal* /ɑ:l/ Engl., 'he said'-'because') can become homophonous in connected speech. We first tested the extent of this phonetic marking of vowel-initial words in a production experiment and found that even in fluent productions, about half of the vowel-initial words are marked with an epenthetic glottal stop. The epenthetic glottal stop is more likely to occur when the preceding word is longer, showing a kind of preboundary lengthening at a phrase-level prosodic boundary. A subsequent perception study (Experiment 2) using a two-alternative forced-choice task with a minimal pair of a glottal stop-initial and a vowel-initial word indicated that listeners are sensitive to the durationally conditioned prosodic context before the test word, and they are more likely to perceive a vowel-initial word when the preceding word is lengthened. An additional eye-tracking study (Experiment 3) using onset-overlap pairs (e.g., *qafla* /ʔafla/ - *afda,* /afda/ → [ʔafda], Engl., 'to trust' - 'chord') showed no early influence of prosodic cues on segmental processing. But a gating experiment (Experiment 4) replicated the prosodic effect observed in Experiment 2. Taken together, our results indicate an interaction between prosodic processing and segmental processing that comes into effect relatively late in speech processing.

Introduction

Spoken words are often produced with different phonetic forms that deviate from their "canonical" forms or the underlying phonological forms. Speech variation, as it must be invariantly mapped on to the intended word, is considered to pose a challenge for spoken-word recognition, especially when the variation creates lexical ambiguity. Understanding how listeners deal with such speech variation has therefore been one of the central issues in developing theories of spoken-word recognition (Gaskell & Marslen-Wilson, 1996, 1997; Gaskell, 2003; Goldinger, 1998; Gow, 2003, 2001; Lahiri & Marslen-Wilson, 1991). A large number of studies have indeed focused on speech variation that occurs at the end of the word, which is often subject to phonological variation such as assimilation (i.e., a sound change where a phoneme becomes similar to a nearby sound in some aspect) and deletion (e.g., Bürki & Gaskell, 2012; Gaskell & Marslen-Wilson, 1996; Gaskell & Snoeren, 2008; Gow, 2003, 2001; Mitterer & Ernestus, 2006). Quite a few studies have also examined speech variation that occurs in the middle of a given word (e.g., vowel reduction to schwa and consonant reduction to flap, Bürki & Gaskell, 2012; Pitt, 2009). These studies have clearly advanced our understanding of speech processing and provided theoretically-grounded insights into how speech variation at the segmental level is dealt with by the listener in spoken-word recognition.

In the present study, we attempt to provide some new insights into spoken-word recognition by focusing on speech variation at the *onset* of the word. Relatively less attention has been paid to the onset of the word (Mitterer & Reinisch, 2015), despite the fact that variation at the word onset may pose a greater challenge for listeners as lexical hypotheses are immediately activated based on the acoustic support for the initial segment (e.g., Marslen-Wilson & Zwitserlood, 1989). This may be mirrored by the fact that phonological processes are less likely to affect the beginning than the end of a word, presumably because speech variation at the onset may be perceptually more detrimental than that at the end (Steriade, 1999). Furthermore, the present study focuses on speech variation that is due to the prosodic context. Previous

* Corresponding author.
E-mail address: tcho@hanyang.ac.kr (T. Cho).

https://doi.org/10.1016/j.jml.2019.104034
Received 17 October 2018; Received in revised form 17 June 2019; Accepted 18 June 2019
Available online 02 July 2019

studies have focused on speech variation due to the *segmental* context, for instance, with deletion of word-final /t/ being more likely if the next word starts with a labial consonant (Mitterer & McQueen, 2009) or a word-medial /t/ being most likely to be deleted when preceded by a nasal (such as /n/) and followed by a weak vowel (e.g., *winter* → *winner*, Pitt, 2009). Speech variation due to phonological assimilation is also driven by the segmental context (e.g., /t/ being pronounced as [p] before another bilabial consonant (such as /m/ or /b/ in *right berry*) Gaskell & Marslen-Wilson, 1996). However, more recent studies have indicated that segmental variation can also be caused by (suprasegmental) prosodic factors, and that spoken words are recognized in reference to prosodic structure—i.e., a structure which determines grouping of words into phrases and distribution of prominence (phrase-level stress) by means of prosodic features such as pitch, duration and amplitude as well as segmental (articulatory/acoustic) strengthening (Cho, McQueen, & Cox, 2007; Kim & Cho, 2013; Kim, Mitterer, & Cho, 2018). In other words, given that phonological or phonetic rules that induce segmental variation are often licensed by prosodic structure (e.g., Cho, Kim, & Kim, 2017; Jun, 1998; Selkirk, 1986), segmental processing is likely modulated by prosodic structural analysis (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Kim & Cho, 2013). This appears to run counter to some current models that assume a strong division in the processing streams for these two aspects of speech (e.g., Giraud & Poeppel, 2012), but a gradually increasing body of studies imply that spoken word recognition involves an interaction of segmental processing with prosodic structural analysis.

The purpose of the present study is therefore to investigate how listeners deal with segmental variation that occurs in a relatively less-studied position—i.e., at the *onset* of a word, and how segmental processing is modulated by or interacts with prosodic structural information available in the speech signal (see Cutler, 2012, for a related review). An excellent case of speech variation that allows us to explore these questions is found in Maltese, a Semitic language, especially with respect to the use of the glottal stop in the language. A glottal stop is a sound in which the airflow is 'stopped' across the glottis (e.g., Ladefoged & Maddieson, 1996). Just as a labial stop (/p/) is produced with the closure of the lips which 'stops' the airflow through the oral constriction, the glottal stop is produced with the abrupt closure (adduction) of the vocal folds which blocks the airflow through the glottis for a short period of time. In Maltese, a glottal stop may be inserted to the beginning of a vowel-word as a phonetic signature of prosodic juncture, but it is also used as a phoneme (or an underlying segment) in the language (see below for more information). Thus, prosodically-conditioned variation of glottalization at the beginning of a vowel-initial word may pose a recognition problem by creating ambiguity, because it is unclear whether the word starts with an underlying glottal stop or is underlyingly vowel-initial. In what follows, we will first discuss the phonological vs. phonetic nature of glottal stops in Maltese, and elaborate on how the phonetic vs. phonological nature of the glottal stop leads to specific questions of the present study.

It is interesting to note that although the glottal stop is used in nearly every language, its function indeed varies across languages. According the UPSID database (Maddieson & Precoda, 1989), about half of the world's languages use it as a phoneme, including the Semitic languages. In Maltese, the glottal stop is a phoneme and can occur in onset and coda position in a syllable, even in consonant clusters with voiced (e.g., *qdart* /ʔdɑrt/ Engl., 'I dared' and *bqajt* /bʔɑjt/ Engl., 'I remained') and unvoiced stops (e.g., *qtates* /ʔtɑtes/, Engl. 'cats' and tqaqpiq / tʔɑʔpɪʔ/, Engl. 'honking of a car horn') (Azzopardi-Alexander & Borg, 1996). English, Dutch, and other languages, on the other hand, use the glottal stop in a non-phonemic way, that is, as a phonetic marker to (optionally) mark the word-boundary of vowel-initial words (e.g., *the eagle,* /ðə#iː.gˢl/ → [ðəʔiː.gˢl]) and as an allophone for an oral stop such as /t/, for example, in a medial position as in *button* or in a final position as in *hit* in some British English accents (e.g., Ladefoged & Johnson, 2014). For English, it has been suggested that the use of glottal marking for vowel-initial words is prosodically conditioned, occurring more often across a prosodic phrase boundary than across a phrase-medial word boundary between a vowel-initial word and the preceding word (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Garellek, 2014; Redi & Shattuck-Hufnagel, 2001). For example, the word *eagle* is more likely to be glottalized in the phrase [[John said]$_{IP}$[eagles are beautiful]$_{IP}$] than in [John said eagles]$_{IP}$. (Note that here 'IP' refers to the Intonational Phrase, a major prosodic phrase in English which is largely defined by certain Intonational properties (known as boundary tones) and accompanying (phrase-final) lengthening towards the end, e.g., Shattuck-Hufnagel & Turk, 1996). Furthermore, it has been suggested that the use of a glottal stop can be phonologically motivated to prevent a hiatus (i.e., a vowel-vowel sequence; in this case when a vowel-initial word is preceded by a vowel-final word). Specifically this was suggested for Dutch (Booij, 1995).

Some languages might use glottal stops in both phonemic and non-phonemic ways. Maltese, a Semitic language with strong influences from Italian and English, provides an interesting case in this regard. Galea (2016) reported that Maltese uses a glottal stop not only as a phoneme but also as a phonetic marker of vowel-initial words[1]. This hence provides a new window on investigating variation in spoken-word recognition. First of all, it allows us to investigate variation at word onset, which is classically considered to be of pivotal importance for word recognition (Marslen-Wilson & Zwitserlood, 1989). If listeners hear the sequence ***il-kelma*** *abjad* /ilkelmɑ#ɑbjɑd/ (where '#' = a word boundary, Engl., 'the word white') produced with an epenthetic glottal stop [ilkelmɑʔɑbjɑt],[2] the intended word *abjad* (Engl., 'white') might be recognized with a significant delay because the epenthetic glottal stop would provide some phonetic (bottom-up) support of the underlying glottal stop as a phoneme, thus creating a potential ambiguity. This is especially so under the assumption of classical models of spoken-word recognition, in that the mental lexicon contains only one phonological form of a given word (e.g., in TRACE, see McClelland & Elman, 1986), which is its canonical form, an assumption that is still part of some current models of word recognition (Kazanina, Bowers, & Idsardi, 2017; Roberts, Wetterlin, & Lahiri, 2013). Under this assumption, when listeners hear the phrase [ilkelmɑ**ʔ**ɑb] in *il-kelma abjad,* the word *abjad* might be deactivated (or its activation substantially attenuated) because the epenthetic glottal stop would cause not only an acoustic mismatch between its initial segment /ɑ/ in the underlying (lexical) representation and the phonetic input [ʔ], but also competition from the word *qabad* (Engl., 'to catch'), which has a glottal stop (represented by the grapheme 'q') in its underlying representation. The word *qabad* is assumed to be more strongly activated than the word *abjad* in this case, given the glottal stop in the acoustic input, eventually deactivating the target word *abjad*[3] (McClelland & Elman, 1986; Norris, 1994).

There are, however, at least three ways in which spoken-word recognition could still achieve relatively efficient recognition of vowel-initial words in Maltese despite an epenthetic glottal stop. First, in line with usage-based approaches to language (Bybee, 2001), it might

[1] Somewhat surprisingly, the words examined by Galea (2016) were underlyingly not vowel-initial but loan verbs from English that all start with an initial geminate (e.g., *to park* → *pparkja, to plug* → *pplugja*). These words usually trigger an epenthetic [i] in connected speech, which is also often found in orthography (*pprakja* /pːɑrkjɑ/ → *ipparkja* [ipːɑrkjɑ]). This epenthetic [i], in turn, triggered epenthetic glottal stops in the data of Galea (2016).

[2] Maltese also applies word-final devoicing, which is why the underlyingly voiced /d/ at the end of *abjad* surfaces as the unvoiced [t].

[3] It could be argued that the problem is artificial, because these orthographically vowel-initial words might have an underlying glottal stop in their canonical phonological representation. However, this possibility is ruled out by phonological processes that apply to vowel-initial words, for instance, with a definite article. The form of the definite article differs between glottal-stop initial words (*il-qattus*, Engl., 'the cat') and vowel-initial words (*l-attur*, Engl., 'the actor'), with an additional vowel only for the glottal-stop initial words.

simply be the case that the mental lexicon of Maltese listeners contains multiple phonetic forms (variants) of vowel-initial words, including forms with and without a glottal stop. Different versions of such episodic models have been proposed, some with the assumption that the listeners store something akin to "grainy spectrograms" (Goldinger, 1998; Pierrehumbert, 2002) while others have proposed that listeners store more abstract variants, similar to narrow phonetic transcriptions or based on an allophonic code (Connine, 2004; McLennan, Luce, & Charles-Luce, 2003; Mitterer, Reinisch, & McQueen, 2018). There is some evidence that listeners indeed store pronunciation variants, so that more than one phonetic form for a given word is stored in the mental lexicon (Brouwer, Mitterer, & Huettig, 2012; Bürki, Ernestus, & Frauenfelder, 2010; Connine & Pinnow, 2006; Pitt, 2009). If this were the case for Maltese vowel-initial words, representations with a glottal stop would be quickly activated by the bottom-up support of an epenthetic glottal stop and would not suffer from the acoustic mismatch described above.

Second, not all phonetic variants of a word may need to be stored in the mental lexicon (Bürki & Gaskell, 2012; Mitterer, Csépe, & Blomert, 2006). Storing variant pronunciations is not necessary if prelexical processing distinguishes the derived (phonetic) forms on the surface from the intended (underlying) forms. If the epenthetic glottal stop in Maltese is prosodically conditioned, as in English (Dilley et al., 1996; Redi & Shattuck-Hufnagel, 2001), listeners might perform a prosodic analysis that bears on the processing of the segmental information. Thus, they would ascribe the glottal stop to a post-lexically driven prosodic function signalling prosodic junctures and therefore not take it into account for lexical access (Kim & Cho, 2013; Kim et al., 2018; Mitterer, Cho, & Kim, 2016).

Third, listeners might use the phonological context in conjunction with the phonetic details to assess whether the glottal stop is underlying or epenthetic. Note that this proposal has an analogue in research on compensation for phonological assimilation, in which both fine phonetic differences (Gow, 2003; Mitterer, Csépe, Honbolygo, & Blomert, 2006) and phonological context (Gaskell & Snoeren, 2008) have been shown to influence compensation for assimilation in spoken-word recognition. Regarding the phonological context, it is conceivable that the epenthetic glottal stop occurs to prevent a hiatus in a V-V context (as suggested by Booij, 1995 for Dutch). Under that assumption, listeners could infer that a glottal stop ([ʔ]) in a V-[ʔ]-V sequence is likely to be epenthetic, but a glottal stop in a C-[ʔ]-V sequence is likely to be an underlying, lexical glottal stop. This inference could additionally be reinforced if there are phonetic differences between epenthetic and underlying glottal stops. Interestingly, there is a well-known phonetic variation of the glottal stop in the languages of the world. Ladefoged and Maddieson (1996) had already noted that the glottal stop is not always produced as a full stop with a closure and a release. Sometimes, it is produced in a reduced fashion with glottalization in which no full glottal closure is achieved (see Appendix B for examples). Mitterer (2018) found that this phonetic variation is used in Maltese to support the phonemic distinction between a word-medial singleton (i.e., phonetically short) and a geminate (e.g., phonetically long) glottal stop. The singleton is likely produced with glottalization but without full closure, whereas a geminate glottal stop in the same word-medial position tends to be produced as a full stop consonant. Despite accompanying duration differences between the singleton and the geminate, listeners make use of such phonetic variation in perception, and they accept a token more often as a geminate if it has a full closure. In a similar fashion, the word-initial underlying glottal stop might also be realized differently from an epenthetic glottal stop. For example, the underlying stop might tend to be produced as a full stop, whereas the epenthetic stop might tend to be produced as the reduced variant— that is, glottalization with no full closure. It is also possible that the underlying glottal stop might be longer than its epenthetic counterpart. The idea that listeners might use such correlates of phonetic strength to distinguish underlying and epenthetic segments would be in line with what Warner and Weber (2001) showed for epenthetic oral stops in German nasal-stop clusters (e.g., the word *Hemd,* Engl., 'shirt', /hemd/ → [hempt]). They found that listeners were more likely to assume an underlying stop if the phonetic evidence for this stop was stronger. Similarly, Maltese listeners might use differences in phonetic strength to distinguish underlying versus epenthetic glottal stops in spoken-word recognition.

Considering all three of those scenarios would expand understanding of the lexical processing of the seemingly ambiguous surface forms of the glottal stop from different theoretical perspectives. It would then eventually provide insights into how listeners deal with the speech variation caused by phonological and prosodic structures. Before exploring those possibilities, however, it is first necessary to know under which circumstances and how often an epenthetic glottal stop actually arises in Maltese. To the best of our knowledge, no systematic, quantitative account of the glottal-stop epenthesis in Maltese has been provided in the literature. We therefore explored the production and distribution of the epenthetic glottal stop in Experiment 1, which used a sentence-production task. Based on the results of that experiment, we then carried out three perception experiments (Experiments 2–4) to assess how listeners might recognize vowel-initial words produced with an epenthetic glottal stop.

Experiment 1

In this experiment, we aimed to elicit vowel-initial and glottal stop-initial words in a sentence generation task without having the speakers focus on the critical words. There are the following questions to be answered: (1) how frequently glottal marking of vowel-initial words occurs in Maltese and whether these glottal markings differ phonetically from underlying glottal stops; (2) to what extent the glottal markings may be related to an avoidance of a hiatus (i.e., a vowel-vowel sequence), which is to be reflected in the extent to which glottal markings occur in relation to whether the word preceding the vowel-initial word ends on a vowel or a consonant; and (3) to what extent glottal markings may be conditioned by prosodic structural context.

We aimed to avoid that speakers would put particular emphasis—or focus—on the critical words because it could influence the likelihood of the glottal stop insertion or the degree of glottalization (e.g., Garellek, 2014). To achieve that, we devised a question–answer game with an implicit contrast between the question and the answer on a word in the test sentence other than the critical word that might bear an epenthetic glottal stop. Fig. 1 provides an example prompt. In the experiment, there were four cartoon characters with which participants were acquainted before the main experiment began. Fig. 1 shows the character with the name 'Matthew'. As in Fig. 1, the letter on the characters' clothing (e.g., 'M') reminded participants of the given name ('Matthew').

Listeners were presented with a written question to which they had to provide a spoken answer based on the picture presented below the question. The change in modality between the question and the answer was necessary: if a participant heard the question in the auditory modality, they would also hear how the questioner produced the test words in the question, which might lead them to alter the way they would pronounce that word (Mitterer & Müsseler, 2013; Pardo et al., 2018; Pardo, 2006).

In Fig. 1, there is a mismatch between the speaker name used in the question ('Daniel') and the one (marked by the letter 'M' for 'Matthew') depicted in the picture. Thus, participants should be answering the question *Does Daniel say the word white in this case?* with the Maltese equivalent of *No,* ***Matthew*** *says the word white in this case*. In this way, there is a contrast on the speaker name, which is likely to lead to a fluent production of the rest of the sentence. The final part *f'dan il-kaz* (Engl. 'in this case') was added to prevent the test word from being influenced by a phrase final-lengthening effect (Turk & Shattuck-Hufnagel, 2007).

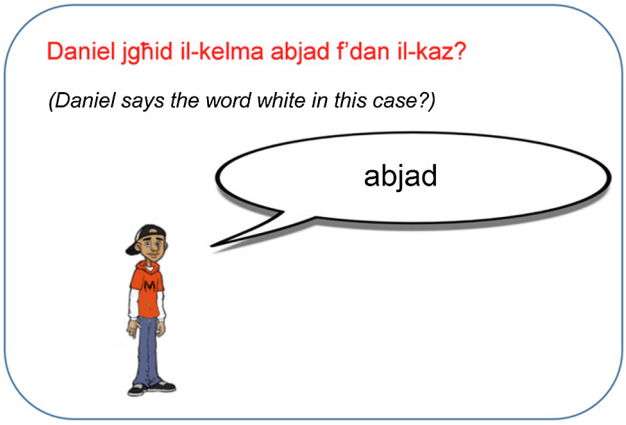

Fig. 1. An example prompt used in Experiment 1 with an added English translation. Participants were asked to answer the question according to the information provided in the picture. Given that the speaker is the cartoon character "Matthew," the correct answer is *Le, Matthew jgħid il-kelma abjad f'dan il-kaz* (Engl. 'No, Matthew said the word 'white' in this case') The critical word *abjad* /ɑbjɔd/ (Engl. 'white') here is vowel-initial and might trigger glottal-stop epenthesis.

The prompt in Fig. 1 leads to an answer in which the critical word (*abjad*) is preceded by a vowel-final word (*il-kelma*), which creates a potential hiatus (*il-kelma abjad*) that might be especially likely to trigger an epenthetic glottal stop. A critical variation to the prompt in Fig. 1 ensured that the critical word could also be preceded by a consonant rather than a vowel. Instead of having one word in the speech balloon as in Fig. 1, the balloon contained two words (*abjad*, Engl., 'white', and *mejda,* Engl., 'table') and the question was *Daniel jgħid il-kliem abjad u mejda f'dan il-kaz* (Engl., 'Daniel says the words white and table in this case'). This leads to the expected answer *Le, Matthew jgħid il-kliem abjad u mejda f'dan il-kaz* (Engl., 'Matthew says the words white and table in this case'), in which the critical word *abjad* is preceded by *il-kliem* /ɪl-kli:m/ (Engl., 'the words'), which ends in an /m/. In this way, we varied whether the target word was preceded by a vowel- or a consonant-final word.

Using this design, we could not experimentally manipulate the size of the potential prosodic boundary between the test word with a potential glottal stop and the preceding context word. In fact, a pilot test (in conjunction with other tasks) had shown that a sentence construction task that attempted such an experimental manipulation could induce a short silence between the two critical words, which would make it difficult, if not impossible, to judge whether there is a glottal gesture between them. In the current task, without an experimental manipulation of boundary size, it is nevertheless possible and even likely that the size of the prosodic boundary could differ across tokens and speakers. We therefore decided to examine possible differences in the size of prosodic boundary within our data to evaluate whether prosody influences the production of an epenthetic glottal stop. (Note that the size of prosodic boundary refers to the boundary strength of whether it, for example, is an Intonational Phrase boundary or a phrase-medial word boundary.) Under the assumption that the glottal-stop epenthesis or the degree of glottalization is positively correlated with the boundary strength (e.g., Dilley et al., 1996; Garellek, 2014), a crucial question here is whether an epenthetic glottal stop is more likely to occur when the preceding word is lengthened. This is based on the phenomenon known as 'preboundary lengthening,' an important correlate of a prosodic juncture (e.g., Cho, 2016; Turk & Shattuck-Hufnagel, 2007), by which a word becomes longer before a larger prosodic boundary than before a smaller one.

Method

Participants

Sixteen students at the University of Malta participated in the study. They were native speakers of Maltese and Maltese English and participated for a small monetary compensation. There were 9 female and 7 male participants, aged 20 to 28.

Stimuli and apparatus

The experiments were performed in a sound-attenuated booth at the Cognitive-Science lab of the University of Malta. The experiments were run on a standard PC using Speechrecorder (Draxler & Jänsch, 2004). Responses were recorded with a Focusrite CM25 large diaphragm condenser microphone connected to a Focusrite 2i2 USB audio interface that did the D/A conversion before storing the files on the computer.

We prepared 134 trials for the experiment. The first four trials were practice runs to familiarize the participants with the procedure. Of the remaining 130 trials, seventy were experimental trials: 35 trials with a vowel-initial test word and 35 trials with a glottal-stop initial test word. For all these words, two prompts were generated, one for the hiatus condition and one for the no-hiatus condition. In both, the word (or words) uttered by the speaker was the same in the question and the depicted scene with the speech balloon. What differed between the question and the depicted scene was the identity of the speaker in the question and the speaker displayed (e.g., 'Daniel' in the question while 'Matthew' was depicted as the speaker). The remaining 60 trials were fillers, consisting of 30 trials in which there was no mismatch between the question sentence and the picture, and 30 trials in which there was a mismatch between the word (the counterpart of the test word) used in the question sentence and the cartoon with the speaker name unchanged. Those should lead to responses such as *Yes*, *Matthew said the words heart and fire in this case,* and *No, Matthew said the word* ***MOUSE*** *and fire in this case*, respectively. An additional 3 prompts were generated for practice purposes and printed out to provide instructions verbally.

Procedure

The experiment started by familiarizing participants with the task using the prompts printed on paper. The research assistant explained the general set-up of the question–answer game. The three printed out prompts contained one example each of a full match (leading to an answer which should not contain a strongly focussed constituent, such as *Yes, Matthew says the word cup*), a mismatch in terms of the speaker name (leading to a contrastive focus on the name as in *Does* ***Nina*** *say the word cup? → No,* ***MATTHEW*** *says the word cup*), and a mismatch in terms of the spoken word (leading to a contrastive focus on another word as in *Does Matthew say the word* ***bottle?*** *→ No, Matthew says the word* ***CUP***). After the familiarization, the main experiment started. The prompts were presented one by one by the Speechrecorder software (Draxler & Jänsch, 2004), starting with the four practice trials and going on with the 130 trials of the main experiment in a randomized order. Speechrecorder was set so that participants had to inspect the display for six seconds before they could start speaking. That is, the Speechrecorder software presented a traffic light icon to the participants to indicate when they could start speaking. It turned from red to green after six seconds, and then the recording started. The recording was ended by the research assistant when the participant had finished speaking. If the participant made an obvious mistake, the trial was repeated by the research assistant.

Two lists were prepared, so that a given test word appeared in one list preceded by a vowel-final word (*il-kelm****a***, Engl., 'the word') and in the other list by a consonant-final word (*il-kliem*, Engl., 'the words').

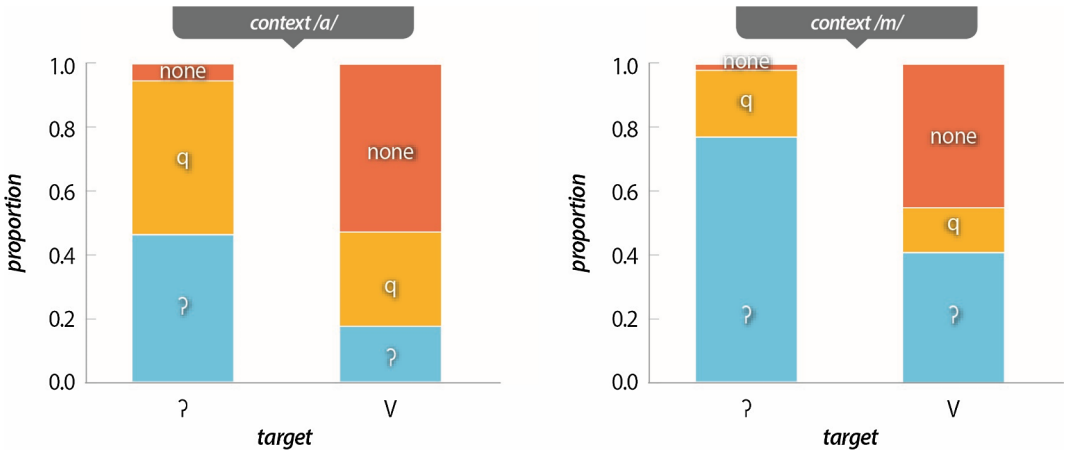

Fig. 2. Proportion of trials in which glottal-stop initial words and vowel-initial words led to a glottal gesture in the production and the form that glottal gesture took (glottal stop "ʔ" versus glottalization "q").

Analysis

The resulting sound files with their intended content were force-aligned using the MAUS web interface (Kisler, Reichel, & Schiel, 2017), which provides a grapheme-to-phoneme transformation for Maltese and then generates a forced alignment of those phonemes. Because MAUS for Maltese does not allow for pronunciation variation, an additional rule set was supplied to the forced-alignment algorithm, indicating that any word-initial vowel might trigger an epenthetic glottal stop and that a word-initial glottal stop might be deleted.

If the algorithm added a pause to the sentence in another position after the initial *Le* (Engl., 'no'), the utterances was discarded as being disfluent. The rest of the resulting alignments were then hand-corrected by the first author. As reported in Mitterer (2018), the forced-alignment algorithm recognizes a glottal stop only when there is a silent period in the speech stream, indicating a full stop of the airflow. This supplies a replicable criterion for when a full glottal stop is present and for the duration of the glottal stop. Because MAUS has a minimal phoneme duration of 30 ms, a glottal stop is only recognized as such if there is a closure of at least 30 ms. In other words, MAUS does not detect glottalization, which thus had to be added during the hand correction. A glottalization was assumed if a clear discontinuity in pitch or amplitude contour occurred (cf. Redi & Shattuck-Hufnagel, 2001). The duration of the glottalization was based on when the pitch or amplitude contour showed the discontinuity (see Appendix B for examples).

To test whether these judgements were reliable, two-hundred utterances for which no glottal stop had been found by the forced-alignment algorithm were additionally transcribed by the second author, blind to the condition and blind to the transcription of the first transcriber. The two-hundred tokens were chosen randomly, with the constraint that half of them were transcribed as vowel-initial by the first transcriber and the other half were transcribed as glottalized. Agreement on the presence versus absence of the glottalization was at 96.5%. Duration judgement for glottalizations correlated at 0.87, indicating that the transcriptions are reliable.

To summarize the coding: first, the forced-alignment algorithm was used to determine whether the speaker produced a full glottal stop with a sustained closure. If that was the case, the forced-aligned duration of that glottal stop was used as a duration measurement. For the rest of the tokens, human raters judged whether there was glottalization and, if so, how long it was, due to the absence of a reliable automatic way to measure the variations speakers can use when producing glottalization.

Results and discussion

Some produced tokens (3.0% of the utterances) were discarded for a disfluency in the critical part of the sentence (i.e., on the target word or the preceding words). Fig. 2 shows the distribution of the phonetic forms for the remaining trials, indicating the proportions in which the glottal stop (epenthetic or underlying) was phonetically realized as a full glottal stop (marked by "**ʔ**" in the figure), as a reduced glottal gesture or glottalization (marked by "**q**" in the figure), or as completely absent (marked by "**none**" in the figure).[4] As shown in the figure, about 50% of the vowel-initial words were produced with a glottal marking (either by "**ʔ**" or "**q**") in both the preceding /a/ and /m/ contexts. In contrast with the phonological motivation for glottal-stop insertion as hiatus prevention, the glottal stop insertion occurred slightly more often in the /m/ context 56.9%) than in the /a/ context (44.2%). This was confirmed by a generalized linear mixed effect model (including a maximal random effect structure for participants and items) that predicted the presence of a glottal marking for vowel-initial words based on context. The regression weight for context was significant ($b = 0.728$, $SE = 0.346$, $z = 2.100$, $p = 0.036$). Thus, the insertion of a glottal marker is not fully motivated by the phonological repair needed to avoid a hiatus. We will provide a tentative explanation for the small effect in the opposite direction below.

Crucially, all vowel-initial words were produced variably, alternating between vowel-initial and glottal stop-initial forms, although the distribution of the variants differed across the test words. Similarly, all speakers produced vowel-initial words with and without glottal stop epenthesis.

It is also important to note that even the underlying glottal stop was produced with both variants. As argued above, it is conceivable that phonetic detail may distinguish underlying and epenthetic glottal stops, which may in turn help listeners infer whether the glottal stop was underlying or epenthetic. Fig. 3 shows the relevant data, comparing the phonetic properties of underlying and epenthetic glottal stops. The left panel shows the likelihood of a full glottal stop versus glottalization, and the right panel shows the duration of the glottal gesture (for a glottal stop and glottalization combined).

As shown in Fig. 3, we found no obvious differences between the underlying and epenthetic stops. Generalized linear mixed-effect models were used to further consider whether there is indeed no evidence to indicate that a given glottal gesture is underlying or epenthetic. The nature of the glottal gesture (underlying vs. epenthetic) was

[4] The patterns shown in Fig. 2 indicate that there are more full glottal stops after /m/ than after /a/. We do not have an objective explanation for the distributional asymmetry due to the preceding context, but it may be at least in part due to how the full glottal stop was defined "only when there is a silent period in the speech stream." The lip closure for /m/ may remain closed after the acoustic nasal murmur (caused by a velum lowering), which may have then contributed to the "silent" period in the speech stream.

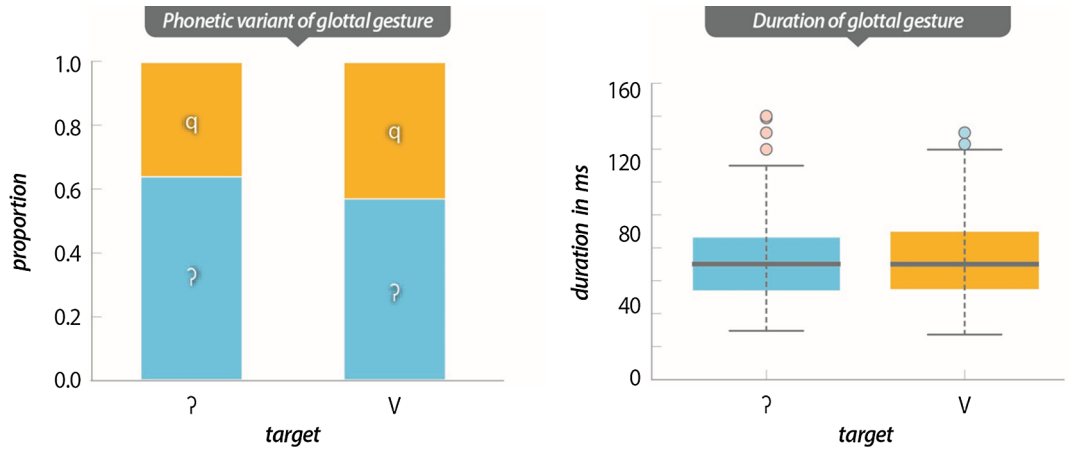

Fig. 3. Comparison of glottal gestures arising from underlying (target = "?") versus epenthetic (target = "V") glottal stops. The left panel shows the proportion of glottal markers realized as full glottal stops "?" versus those realized as glottalization "q". The right panel shows a boxplot of the duration of the glottal gestures for epenthetic versus underlying cases.

used as the dependent variable, with the underlying stop mapped onto 1. The independent variables were the duration of the glottal gesture and the type of gesture (with a full glottal stop coded as 1 and glottalization coded as 0). The model had a random effect for speaker and a random slope for the type of glottal gesture (full stop vs glottalization) over speaker. Duration as a random slope was not included because it led to convergence failure. A random effect for item was not included either because it does not make sense in this case. That is, given that the dependent variable is whether the word is vowel-initial or glottal stop-initial, it will reflect a random effect for item. Even though full glottal stops were (unsurprisingly) longer than the reduced glottalizations (78 ms vs 64 ms), the two predictors are not strongly collinear with a point biserial correlation of 0.27.

Results from the generalized linear mixed-effect model are summarized in Table 1. If the underlying glottal stops are phonetically stronger (i.e., more occurrence of full glottal stops with a sustained closure or longer duration), the regression weights should be positive. However, as Table 1 indicates, the regressions weights are negative and fail to reach significance. This indicates that phonetic classification and duration do not hold as reliable cues to whether a glottal gesture is underlying or epenthetic.

While it remains possible that a significant difference might be found with a larger sample, such a difference is unlikely to be useful for the listener for the task of spoken-word recognition. An interesting analogy here is provided by incomplete devoicing in German, which is only reliably detectable with a considerably large sample (Roettger, Winter, Grawunder, Kirby, & Grice, 2014). Our data base is comparable to that of Roettger et al. (2014) who also used 16 participants and 48 and 96 items in their first and second experiment, respectively; the current study has 16 participants and 70 items. Even though Roettger et al. (2014) found a significant effect of underlying voicing on acoustic cues, they also deemed it likely that "… the acoustic cues found in the neutralized position have no functional utility and are not reliably used in regular communication to differentiate between minimal pairs" (p. 22). In line with assumption, there is no evidence to suggest that listeners can use the cues left by incomplete devoicing to constrain lexical access.

Next, we analysed whether the presence or absence of an epenthetic glottal gesture for a vowel-initial word was more likely to occur when the preceding word was lengthened. To this end, we first determined the *relative duration* of the preceding word in a given sentence after accounting for influences of articulation rate (measured as the number of syllables per second, based on the duration of the other words of the sentence) and the nature of the preceding word (i.e., *il-kelma* or *il-kliem,* with the expectation that *il-kelma* is likely to be longer as it has one syllable more than *il-kliem*). This was achieved with a linear-mixed effect model with these two factors and a random effect for speaker, which included a random slope for the nature of the preceding word, and preceding word duration as dependent variable. The result (see Table 2) shows the expected effects: The preceding word is shorter when the articulation rate (measures as syllables per second) is higher and it is shorter when it is the two-syllable *il-kliem* rather than the three syllable *il-kelma*. This residual of the model represents whether the preceding word was relatively short or long. This measure of relative duration was then used as a fixed effect in a generalized linear mixed effect model that predicted the presence or absence of a glottal stop with random effects for participant and item, including random slopes for the duration of the preceding word. This model revealed that, the longer the preceding word, the more likely it is that there is an epenthetic glottal stop (b = 9.423 (2.848), z = 3.308, p < 0.001). Given the intercept (0.099) and the standard deviation of the relative duration (0.047 s), this means that the likelihood of an epenthetic glottal stop is 41.3% when the preceding duration is one standard deviation below the mean but 63% when the preceding duration is one standard deviation above the mean. Thus, when the prosodic juncture between the two words is relatively large (as reflected in the longer duration of the preceding word), a glottal gesture is more likely to be inserted than when it is relatively small.[5]

Table 1
Results from the generalized linear mixed-effect model trying to predict the nature of a glottal stop (underlying or epenthetic) from the glottal gesture type (a full stop versus glottalization) and duration.

	b (SE)	Z	p
Intercept	0.902 (0.154)	5.843	< 0.001
Glottal gesture = full	−0.411 (0.232)	−1.772	0.078
Duration in ms	−0.110 (0.083)	−1.338	0.181

Table 2
Duration of the preceding word before a vowel-initial word, as predicted by the estimated articulation rate and whether the word was *il-kelma* or *il-kliem.*

	b (SE)	t (df)	p
Intercept	456.479	49.115 (14.185)	< .001
Articulation rate	−14.526	−2.983 (493.445)	.003
word *il-kliem*	−40.775	−4.833 (15.002)	< .001

[5] This result may also explain the somewhat unexpected small preference for

To summarize, the results of Experiment 1 show that the glottal marking of a vowel initial word is frequent in Maltese, with about half of such words produced with a full glottal stop or glottalization. Our analysis of the distribution of the glottal markers shows that their occurrence is not motivated by a phonological repair of a hiatus and that the phonetic properties of the glottal marker (as reflected in the duration and type of glottal gesture) are similar to those of underlying glottal stops. However, the prosodic boundary had a small but significant effect such that the glottal markers were more likely to occur when the speaker lengthened the preceding word (i.e., preboundary lengthening). Next, we wondered how listeners would deal with a speech signal containing a glottal gesture when they were perceiving vowel-initial vs. glottal stop-initial words and whether the observed prosodic contextual effect on the realization of the glottal marker is indeed available to the listener and used in speech perception. To answer those questions, we conducted Experiment 2.

Experiment 2

In this experiment, participants performed a two-alternative forced choice (2AFC) task between a glottal stop-initial word and a vowel-initial word in a minimal pair (*għam* /ɑ:m/ *-qam* /ʔɑ:m/, Engl., 'he swam' - 'he woke up'). A phonetic continuum between unglottalized and glottalized versions was generated based on a natural recording of the vowel-initial word *għam*, lowering the pitch and amplitude on the initial vowel (two cues typically associated with glottalization, cf. Redi & Shattuck-Hufnagel, 2001) using the PSOLA algorithm in Praat (Boersma, 2001). The resulting stimuli were spliced into a sentence in which the word preceding the critical word was slowed down or not, creating two prosodic conditions (± preboundary lengthening). Given that preboundary lengthening is generally correlated with boundary strength, those conditions provide another example of an interaction between prosodic and segmental processing in speech perception (Kim & Cho, 2013; Mitterer et al., 2016, who also used a 2AFC task). In other words, segmental perception varies as a function of perceived boundary strength. We therefore hypothesized that listeners would be more likely to attribute the phonetic evidence for glottalization in the speech signal to a prosodically conditioned epenthetic glottal stop than to an underlying phoneme in the lengthened condition than in the non-lengthened condition. In other words, if our hypothesis is correct, listeners will perceive glottalized stimuli as vowel-initial words more often in the lengthened (+preboundary lengthening) condition than in the non-lengthened (-preboundary lengthening) condition.

Method

Participants

12 students at the University of Malta participated in this experiment. They were native speakers of Maltese and Maltese English and participated for a small monetary compensation. There were 9 female and 3 male participants, aged 20 to 26.

Apparatus

This experiment was performed in a sound-attenuated booth at the Cognitive-Science lab of the University of Malta. Experiments were run on a standard PC using PsychoPy (version 1.84, Peirce, 2007). Sounds were presented using Logitech Z 150 speakers positioned on the right and left of a 22-inch monitor.

Stimuli

An adult male speaker of Maltese produced multiple renditions of the sentences *tikteb il-kliem* (*għam/qam*) *u nar* (Engl., 'She writes the words (he swam/he woke up) and fire', 3.35 vs 15.75 occurrences per million words, respectively, according to the MLRS corpus, Gatt & Čéplö, 2013). From one utterance, the parts preceding and following the critical word were spliced out to form a sentence frame. The preceding part was manipulated with PSOLA in Praat to generate two versions, one that had the same timing as the original, fluent utterance which was not produced with preboundary lengthening and one that was manipulated to contain preboundary lengthening. This could be easily achieved by simply lengthening the utterance and use the original and lengthened utterances. But with this simple procedure alone, the two stimuli would not only differ in duration but also in the amount of modification they had undergone. To avoid this confound, we applied the following procedure: The voiced part of the word *kliem* was lengthened, so that it was 25 ms longer than the original. Then, this lengthened version was manipulated further, once using a lengthening factor and once using a shortening factor for the duration manipulation. Thus, both types of carrier sentences were re-synthesized twice with the PSOLA algorithm applying a duration manipulation. One of these had the same duration as the original sentence, which constituted the [-preboundary lengthening] condition. In the other stimulus, the word *kliem* preceding the target word (*għam/qam*) was 55 ms longer than in the original, and this was the [+preboundary lengthening] condition.

For the target continuum, the initial 50 ms of the vowel-initial target word, originally produced without any phonetic evidence of glottalization, were pitch- and amplitude-manipulated to mimic the typical properties of glottalized vowels (Redi & Shattuck-Hufnagel, 2001): reduction in amplitude and the F0 (pitch) dip. Pitch was lowered from 100 to 60 Hz, and amplitude was lowered from 100% of the original to 50% of the original in 6 steps. (It should be noted that in addition to the F0 dip (pitch lowering) and reduction in amplitude, other observable phonetic cues to glottalization include irregularity in the spacing of pitch periods and long uninterrupted decay times. We chose the F0 dip and amplitude cues not only because they are typical cues to glottalization, but also because they are quantitative measures that can be reliably manipulated along a phonetic continuum.) Fig. 4 shows the two endpoint stimuli and one mid-point stimulus on the continuum: Step 1 has the largest amplitude and the highest F0 (i.e., no glottalization), and Step 6 has the most reduced amplitude and the lowest F0 (i.e., strongest glottalization) at the vowel onset. The primary purpose of using the phonetic continuum is to test the contextual (preceding) lengthening effect on the perception of the glottalization across various stimuli, not to test whether the perception between an underlying glottal stop and an epenthetic one can be shifted as a function of glottalization strength. However, the continuum should influence whether the participants perceive a glottal gesture. Given that perceiving a glottal gesture in the first place is a prerequisite to attributing the glottal stop to an underlying phoneme, we still expect more underlying glottal-stop responses as the glottalization cues become stronger.

Procedure

All instructions were given via the computer screen as part of the experiment. Participants were told that they would hear sentences such as *tikteb il-kliem* ***X*** *u nar* (Engl., 'She writes the words **X** and fire') and that the word in the position of the "X" mark could be *għam* (/ɑ:m/) or *qam* (/ʔɑ:m/). They were told that these two words would be presented on the left and right side of the computer screen and that their task was to indicate which of the two words sounded more like the one they heard in the sentence by pressing the corresponding left or right arrow key on the computer keyboard.

Participants heard each of the 12 stimuli (± preboundary

(*footnote continued*)
insertion of a glottal stop after *il-kliem* (Eng., 'the words') than after *il-kelma* (Engl., 'the word'). We only have a speculative explanation to offer: After *il-kliem*, participants have to plan a slightly longer utterance, which may lead them to introduce a glottal stop to give them more planning time (similar to adding an optional function word to a sentence to gain planning time, see Ferreira & Dell, 2000).

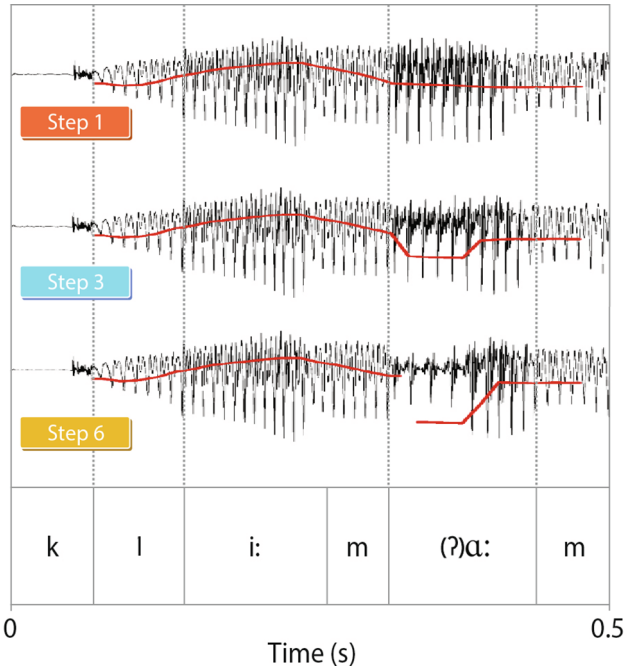

Fig. 4. Example stimuli from the zero-to-glottal stop continuum used in Experiment 2, as reflected in the waveforms and pitch contours (red lines). The first step (Step 1) has no cues for glottalization (with the largest amplitude and no F0 dip), and the last step (Step 6) contains the strongest cues, with the lowest amplitude and largest F0 dip.

lengthening with 6 levels of glottalization cues) twelve times. (Participants also heard six more stimuli from a continuum with a pitch accent on the target word, but those data are not reported here.) The stimuli were presented in twelve blocks, and the order of the stimuli was randomized within each block.

Results and discussion

Fig. 5 shows the results from the 2AFC task after the rejection of 12 trials (0.69%) because of slow reaction times (> 4s). The results showed that participants perceived an underlying, lexical glottal stop

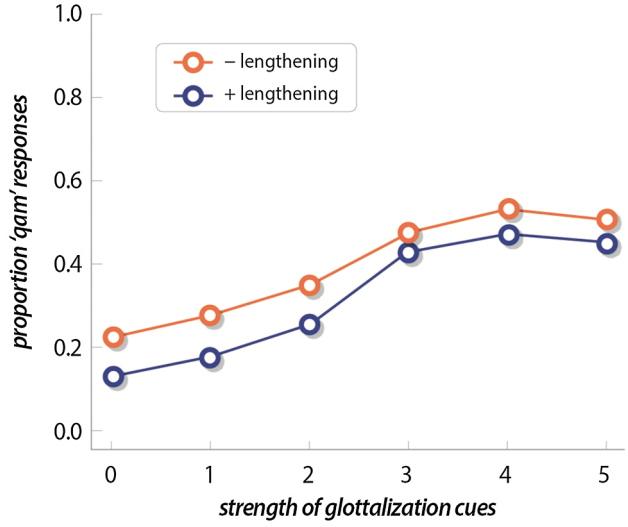

Fig. 5. Mean proportions of *qam* responses in which participants perceived an underlying glottal stop for the two lengthening conditions (± preboundary lengthening) over the continuum. ('0' refers to Step 1, and '5' refers to Step 6 from Fig. 4).

Table 3
Results from the generalized linear mixed effect model for the likelihood of *qam*-responses during Experiment 2. Predictors were contrast coded (strength of glottalization from −2.5 to 2.5 and ± preboundary lengthening as ± 0.5).

	Estimate	Std. Error	z	p
(Intercept)	−0.684	0.348	−1.963	0.050
Strength of glottalization	0.389	0.080	4.846	< .0001
Preboundary lengthening	−0.516	0.122	−4.249	< 0.001
Strength of glottalization: Preboundary lengthening	0.132	0.073	1.801	0.072

(i.e., *qam* (/ʔɑ:m/) more often when there was no boundary cue (i.e., –preboundary lengthening, 40% *qam* responses) than when there was a boundary cue (i.e., +preboundary lengthening, 32% *qam* responses). They also perceived an underlying glottal stop more often at higher steps (when the stimuli contained stronger glottalization cues) than at lower steps, which probably reflects that listeners are more likely to perceive a glottal gesture as the cues for such a gesture become stronger. Those patterns were supported by a generalized mixed-effect model with a logistic linking function (see Table 3). The likelihood of a *qam* response was the dependent variable, and the boundary condition (± preboundary lengthening) and glottalization cue strength (6 levels) were the independent variables. We used the independent variables as contrast-coded fixed factors (± preboundary lengthening → ± 0.5, six levels of strength of glottalization → [−2.5, −1.5, −0.5, 0.5, 1.5, 2.5]) and included their interactions. A random effect for participants was used with a full random-effect structure, save for correlations between random effects because the inclusion of those is likely to lead to convergence failure.

The results indicate that in Maltese, listeners' perceptions are influenced by the prosodic conditioning of the glottal-stop insertion (the effect of preboundary lengthening). Listeners are *less* likely to perceive a word as having an underlying glottal stop when there is an indication of a prosodic boundary, which might in turn increase the likelihood that the speaker intended a vowel-initial word that received an initial glottal marking. In other words, the results suggest that listeners are *more* likely to attribute the glottalization to the boundary-induced glottal marking in the presence rather than in the absence of preboundary lengthening. This thus provides another example of an interaction between prosodic and segmental processing, as previously reported in Kim and Cho (2013), Mitterer et al. (2016) and Steffman (2019). Based on the finding that the categorical perception function of voiced versus voiceless stops in American English is modulated by lengthening the preceding context, Kim and Cho (2013) argued that segmental processing occurs with reference to prosodic structure. A more recent study of this topic (Kim et al., 2018) further noted that this interaction might arise only late in processing. Kim et al. (2018) tested the time course of the interaction between prosodic and segmental processing by using an eye-tracking paradigm. Although the data suggested that this interaction might arise early, only late effects were statistically significant. Their results therefore raise the possibility that the initial analysis of segmental content in lexical processing is modular and not influenced by prosodic processing, which comes into play only at a later stage of speech processing. In Experiment 3, we also explored the time course of the interaction between prosodic and segmental processing by using an eye-tracking paradigm. We specifically tested whether the effect of the boundary condition (as signalled by preboundary lengthening) indeed comes later than the processing of segmental information for glottalization, which provides phonetic support for both an underlying and epenthetic glottal stop.

Experiment 3

Consider a case in which a Maltese listener hears the fragment *jifhem qab…* [jifhemʔɑb…], (Engl., 'he understands *qab…*'). Given that

listeners attempt lexical access as early as possible based on the available acoustic information (Allopenna, Magnuson, & Tanenhaus, 1998), listeners might activate all kinds of Maltese words that start with *qab* (e.g., *qabel* 'before', *qabras* 'to have a great time', *qabad* 'to catch'). Under the common assumption of the lateral inhibition of words in spoken-word recognition (e.g., McClelland & Elman, 1986) and that vowel-initial words are represented in the mental lexicon as such (i.e., vowel-initial), this cohort of words should strongly inhibit words that do not start with a glottal stop such as *abjad* (Engl., 'white'). However, as the results of Experiment 1 showed, a phrase such as *jifhem abjad* might actually give rise to a competing fragment *jifhem [ʔ]ab…* due to the frequent occurrence of glottalization or an epenthetic glottal stop. In such a case, it is possible that a listener's inference (i.e., the glottal stop [ʔ] is epenthetic) increases when there is preboundary lengthening on the preceding word, a prosodic context in which an epenthetic glottal gesture is expected (see Experiments 1 and 2).

In Experiment 3, we explored that possibility using an eye-tracking task with a visual world paradigm in which the word-onsets of the targets are similar, except for the presence or absence of an underlying glottal stop (e.g., *qabad* and *abjad*). We call them *pseudo onset-overlap pairs*. Given the phonological constraints of such target words, we could not use pictures as the targets. Instead, we asked participants to click on printed words (McQueen & Viebahn, 2007). The participants heard sentences such as *Nina tifhem* ([ʔ])*abjad* (Engl., 'Nina understands white') and saw the printed words *abjad*, *qabad*, and two unrelated distractors on the screen. Their task was to click on the object of the verb *jifhem* (Engl. '(he) understands'), the vowel-initial word *abjad* in this case. There were two independent variables: target type and prosody cue. The target type refers to the presence or absence of an underlying glottal stop at the word onset (i.e., does the target word start with an underlying glottal stop or is it vowel-initial?). The prosody cue refers to the presence or absence of preboundary lengthening (i.e., is the syllable preceding the target lengthened, thereby providing a cue for a prosodic boundary?). The dependent variable was the amount of fixation on the target versus the competitor, which would indicate the extent to which the target is preferred over the competitor.

This design can address two questions. First, it can address whether the influence of prosody (i.e., preboundary lengthening) on segmental processing is fast enough to influence the initial evaluation of a glottal stop. If that is the case, participants who hear a phrase such as *Nina tifhem* ([ʔ])*ab…* should be more inclined to look at the vowel-initial word in the lengthened condition (i.e., when the preceding word *tifhem* is lengthened, which indicates a prosodic boundary larger than a typical word boundary). Statistically, that would be expressed as an interaction between the target type (vowel-initial vs. glottal-stop initial) and the prosody cue (present vs. absent). For vowel-initial words, adding preboundary lengthening should increase the likelihood of a look toward the (vowel-initial) target word; whereas, for glottal-stop initial words, preboundary lengthening should lead to more looks to the vowel-initial competitor and fewer looks to the glottal-stop initial target, giving rise to an interaction.

In contrast, if the influence of prosody comes into effect relatively late, no such effect is likely to be observed because the remaining part of the phrase will quickly disambiguate the target. That is, when participants hear the later part of *Nina tifhem* ([ʔ])*abad* (Engl. 'Nina understands catch'), the segments (*…bad*) following the initial vowel will have already ruled out the possibility that the speaker intended to say the vowel initial word that contained *abjad*.

Second, we expected a general preference for words with an underlying glottal stop, which would appear as a main effect for target type. This prediction is based on two ideas. First, given that only half the vowel-initial words were marked with a glottal gesture in Experiment 1, upon hearing a glottal gesture (in contrast with the lack of a glottal gesture in some vowel-initial words), listeners should prefer the interpretation that the intended word has an underlying glottal stop. Second, several previous findings in spoken-word recognition indicated that listeners generally prefer to attribute information in a speech signal to an underlying phoneme rather than to a phonological process. For instance, Ohala and Ohala (1995) found that listeners preferred to interpret a nasalized vowel as an underlying nasal vowel rather than assuming that it was an oral vowel that had been nasalized from a neighbouring nasal consonant. In our own previous work (Kim et al., 2018; Mitterer, Kim, & Cho, 2013), we used Korean phonological alterations (post-obstruent tensing and consonantal place assimilation, respectively), which were largely categorical (i.e., the derived form is not distinguishable from an intended form). For instance, participants in Mitterer et al. (2013) heard words that ended in a clear velar stop but were underlyingly labial due to a labial-to-velar assimilation. When participants heard those words, they had a strong preference for targets with an underlying velar even in the assimilatory environment, and they required a sentence context to look at the intended word, which had an underlying labial realized as a velar at the surface. It is therefore reasonable to predict that listeners would be faithful to bottom-up acoustic information in segmental processing, at least at an early stage of lexical processing, irrespective of when in the process a prosodic analysis comes into effect.

Method

Participants

Forty-one students at the University of Malta participated in this experiment. They were native speakers of Maltese and Maltese English and participated for a small monetary compensation. There were 14 female and 12 male participants, aged 18 to 27. Additionally, two staff members at the University of Malta also participated, but because they were much older than the other participants (> 40), their data were not analysed.

Apparatus

Experiments were performed in a sound-attenuated booth at the Cognitive-Science lab at the University of Malta. They were run on a standard PC using Experiment Builder, and eye movements were tracked with an Eyelink 1000 eye-tracker in desktop mode at a frequency of 500 Hz.

Stimuli

An adult female speaker of Maltese produced multiple renditions of sentences such as *[Matthew|Daniel|Mary|Jenny] [j|t]ifhem TARGET* (Engl. '[Matthew|Daniel|Mary|Jenny] understands TARGET'). The alteration between *jifhem* and *tifhem* was necessary because Maltese uses different forms for the masculine and feminine third person (i.e., *Matthew jifhem* vs. *Jenny tifhem*).

We used 48 pseudo onset-overlap pairs of vowel-initial words and glottal-stop initial words (see Appendix A), and recorded each one twice. Materials for an additional 120 filler trials were generated, with the target words of those trials using neither a vowel- nor glottal stop-initial. Some target words were also recorded twice, with two different visual prompts for the speaker, one with a colon before the target word and one without. This was intended to induce different prosodic phrasings (i.e., *Matthew jihfem TARGET* → [Matthew jifhem TARGET] vs. *Matthew jihfem: TARGET* [[Matthew jifhem] TARGET].

To control the strength of the cues for a glottal stop in the critical pseudo onset-overlap pairs (such as *qabad-abjad*), we used cross-splicing. Each of the two words of such pair had been recorded twice. From these four recorded words (two tokens of a vowel-initial word and two tokens of a glottal stop-initial word), we selected one that was identified as containing a clear glottalization. This selection could either be the vowel-initial or the glottal stop-initial word. The part of glottalization in the selected token was then spliced out of the utterance and spliced into the other recorded token of the same word which was used as a stimulus. The same part of glottalization was also spliced into one of the two recorded tokens of the other word which formed a pair. In

this way, the two members of a stimulus pair contained the same glottal stop and also were both cross-spliced.

To generate different versions of the precursors, we calculated the difference in duration of the last syllable of *jifhem* that occurred with the four different proper names. The ratio ranged from 1.6 to 1.8, and the stimuli used here were generated to have a lengthening close to the maximum of this distribution. Thereby, the design of this experiment maximizes our chance of finding an effect while still presenting stimuli within the normal range of prosodic lengthening. Lengthening was achieved using PSOLA on the basis of utterances that contained a word boundary (smaller than a prosodic phrase boundary) between the target word and its preceding word (*jifhem*). The first duration point was set at the onset of *jifhem* with a new/old duration ratio of 1, meaning that the first part of the utterance will not be changed in terms of duration. A second duration point was then set at the onset of the last syllable (i.e., *hem* in *jifhem*). Again, an intermediate stimulus was generated, with a duration ratio of 1.35 (i.e., a 35% lengthening of the final syllable). The PSOLA syntheses for these stimuli were generated and then subjected to another PSOLA manipulation. For one stimulus, the second duration point was set at 1 over 1.35, the inverse of the original manipulation, to generate a stimulus with the original duration ratios that has undergone the same amount of manipulation as the stimulus with lengthening. The stimulus with lengthening was also generated by setting the second duration point to 1.35, leading to a lengthening of 1.82 ($=1.35^2$). The two versions of the four precursors (due to the four different proper names used in the experiment) were then concatenated with the target words, creating two preboundary lengthening conditions: ± preboundary lengthening.

For the visual display, unrelated distractor words were added to the critical pseudo onset-overlap pairs so that there were four words on the screen. Of the fillers, forty were other onset-overlap pairs (e.g. *ballun-baliena*, Engl., 'ball'-'whale') plus an unrelated distractor on the screen to prevent participants from assuming that, if there are two phonologically similar words on the screen, one of them is likely to be the target. For the remaining 80 filler trials, a vowel-initial word or a glottal stop-initial word was used as one of the distractors, again, to discourage participants from assuming that any vowel-initial or glottal stop-initial word on the screen was likely to be the target.

Procedure

Participants first read an instruction that familiarized them with the visual-word paradigm. They were instructed to click on the word that "was understood", that is, the object (TARGET) of the sentence *[Matthew|Daniel|Mary|Jenny] [j|t]ifhem TARGET*. After they read the instructions, the eye-tracker was calibrated using a nine-point calibration, and then the main experiment began.

Each participant completed 168 trials (48 experimental trials and 120 fillers). The experiment started with 3 filler trials. A different random order was generated for each participant, with the following constraints: each critical pair was presented once, and the condition for that pair was counterbalanced over participants. Moreover, the target and competitor positions were counterbalanced for each participant, so that each of the twelve possible combinations of the target and competitor positions occurred once in each experimental condition and ten times in the 120 filler trials. We did that to ensure that participants' preference to start scanning at the upper left corner of the screen did not influence the results. After every 50 trials, participants were told how many trials they had completed and had the opportunity to take a short break.

Results and discussion

Our results indicate that participants clicked on the intended word in about 97% of the cases, with very little spread between the experimental conditions (min: 96.8%, max: 97.8%). There were, however, some numerical differences in click latencies, with slightly longer RTs for vowel-initial words (+boundary: 1408 ms, -boundary: 1415 ms) than glottal-stop initial words (+boundary: 1398 ms, -boundary: 1361 ms). A linear mixed-effect model was therefore used with the natural logarithm of the reaction time as the dependent variable, the contrast-coded fixed factor preboundary lengthening (± 0.5) and target type (−0.5: vowel initial, +0.5: glottal-stop initial) plus their interactions, and random factors for participant and item. The random-effect structure was maximal (but note that no random slope of target type over item was used because the items are either vowel initial or glottal-stop initial). As Table 4 shows, there were no statistically significant effects.[6]

Table 4
Results of the linear mixed-effect model for log(rt) in Experiment 3. The results show no significant influence of the independent variables on click latencies.

	b (SE)	t (df)	p
Intercept	7.193 (0.038)	187.438 (26.8)	< .001
Preboundary lengthening	0.011 (0.016)	0.681 (19.9)	0.504
Target type	−0.023 (0.022)	−1.059 (45.6)	0.295
Preboundary lengthening : target type	0.028 (0.03)	0.931 (1045)	0.352

Fig. 6 shows the eye-tracking data, which indicate a small preference for the glottal-stop initial items in both boundary conditions, with no clear difference due to preboundary lengthening. To analyse the data, we used a measure of target preference, which was the difference in the fixation on the targets (versus the competitors) in the time window 200–600 ms after target onset, transformed into logOdds. We used this size of the analysis window to focus on early effects. Given that it usually takes 150–200 ms for the speech signal to influence eye-movements, the lower (left) bound of the window was set at 200 ms. The upper (right) bound (i.e., 600 ms) was chosen in such a way that the resulting 400 ms window would allow for some aggregation to reflect the full extent of early processing. Luck (2005) noted that the earliest effects in time-varying data indicate the fastest responses by the fastest participants, thereby motivating a relatively longer window to capture all available evidence of early processing.

This dependent measure was used in a mixed-effect model similar to the one used for reaction time. The model had the contrast-coded fixed factors of target type, preboundary lengthening, and their interaction as predictors, again with a full random-effect structure. As a control variable, the target word frequency (based on the Maltese Language Resource server corpus, v3.0, see Gatt & Čéplö, 2013) was used because the log per million frequency of the vowel-initial words (1.63) was slightly higher than that of the glottal-stop initial words (1.06). Note, however, that this was a small effect (d = 0.23, SD = 2.63). The results of this analysis are given in Table 5, which indicates no significant effect of target type and prosody (boundary) cue and, critically, no interaction between the two.[7]

Given the absence of any significant effect, the question arises, whether the experiment was not sensitive enough. Therefore, we compared the current experiment with other experiments that also investigated variant recognition using eye-tracking with a generally

[6] We used the logarithm of the reaction time so that their distribution would be closer to normal. A reviewer suggested that an effect might be found in especially long reaction times, which may less impact if the log is used. We therefore also ran the analysis on raw reaction times, with a similar result (Intercept: 1394; Preboundary lengthening: b = −14 (18), t(89) = −0.771, p = 0.443, Target type: b = −32 (32), t(77) = −1.018 , p =0.312; Interaction: b = 24 (36), t(89) = 0.658, p = 0.512).

[7] Based on the suggestion of a reviewer, we also did an exploratory analysis of a later time window, since Fig. 6 appears to suggest an interaction. Using 800-1000ms time window, the interaction was not significant, either (b = −0.524 (SE = 0.282), t(1708) = −1.855, p = 0.0637). Note, however, that focusing time windows on where effects seem strongest is problematic (e.g., Luck, 2005).

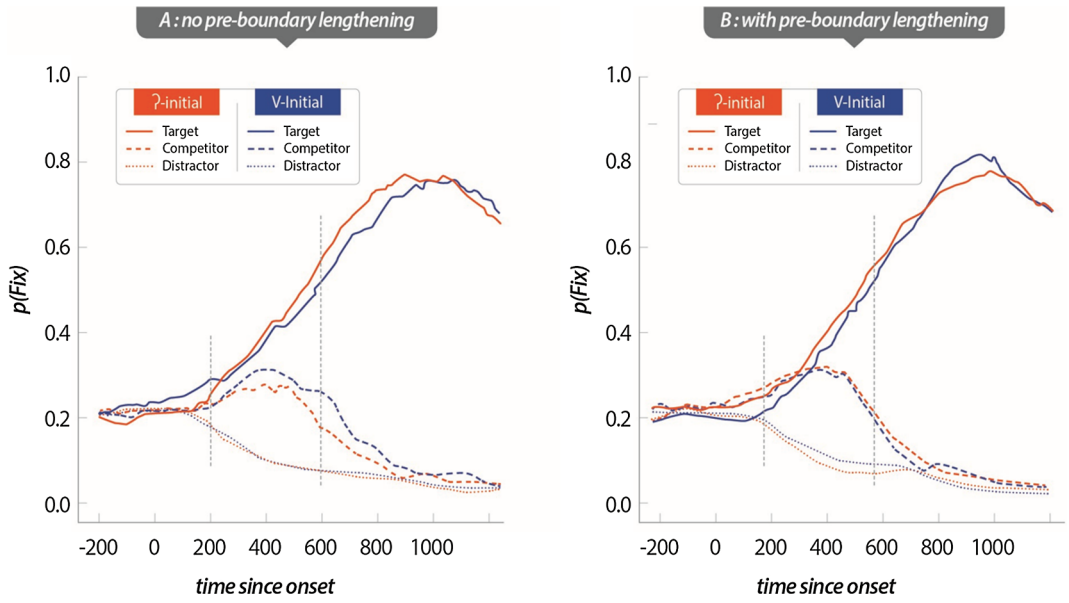

Fig. 6. Fixation proportions in Experiment 3 for the target, competitors, and distractors, depending on the boundary cues of preboundary lengthening (different panels) and target word (line colour). The vertical dashed lines indicate the analysis time window.

Table 5
Results from the generalized linear mixed effect model for the eye-tracking target preference 200–600 ms after target onset in Experiment 3 (see text for details of the model).

	b (SE)	t (df)	p
Intercept	0.761 (0.113)	6.762 (97)	< .001
Preboundary lengthening	−0.335 (0.218)	−1.54 (37)	0.132
Target type	0.267 (0.2)	1.34 (95)	0.183
Target frequency	−0.056 (0.038)	−1.469 (97)	0.145
Preboundary lengthening : target type	−0.002 (0.394)	−0.006 (95)	0.995

smaller number of items (Kim et al., 2018; Mitterer & Reinisch, 2015; Mitterer et al., 2013). We re-analysed the data from these experiments using the same measure as in the current experiment (i.e., target over competitor preference in a 400 ms window after the onset of the mismatching information), and found significant mean differences that ranged from 1.3 to 1.9 logit units (all *ps* < 0.0001).[8] With the same method, our data in the current experiment did not show any significant difference with standard errors ranging from 0.2 to 0.4 logit units (see Table 5). Moreover, we analysed the power of the experiment using the method proposed by Westfall, Kenny, and Judd (2014). Using the variance portioning as observed in our experiment, our design with 41 participants and 96 items has adequate power (0.8) to find an effect size of 0.31. In the experiments cited above, effect sizes range from 0.26 to 0.51, meaning that we have adequate power to find the typical effect of mismatch caused by a phonological variant.

We hence observe that there is no robust preference for activating glottal stop-initial items over vowel-initial items when hearing a glottal stop. This is a somewhat surprising finding, given that eye-tracking experiments have consistently found a strong preference for competitors whose onsets overlap with those of the targets over those with an onset mismatch (Allopenna et al., 1998; Brouwer et al., 2012; Kim et al., 2018; Mitterer & Reinisch, 2015; Mitterer et al., 2013). It is especially noteworthy that vowel-initial words turned out to be strong competitors for glottal-stop initial words. The shape of the competitor curves in Fig. 6 looks much like those observed with a cohort competitor (i.e., competitors that are identical at the onset with the target word), even though the vowel-initial words have—at least underlyingly—a different onset phoneme. The results of the current eye-tracking experiment therefore suggest that the phonetic evidence for a glottal stop does not lead to strong deactivation of a vowel-initial target. This finding can be explained by assuming that Maltese listeners store a phonetic pronunciation variant—or multiple acoustic variants—of vowel-initial words with a glottal stop in their mental lexicon. The phonetic form (a V-initial word with glottalization) that is matched with such a representation or representations would be accessed relatively quickly using bottom-up input, which in turn would explain why vowel-initial words do not suffer from huge recognition costs when they appear with an epenthetic glottal stop and why they function as strong competitors for glottal stop-initial words.

The absence of an interaction between target type and prosody (boundary) cue indicates that the influence of preboundary lengthening does not come into effect early enough to influence the initial assessment of the glottal stop, which contrasts with the results of Experiment 2. We have two competing explanations. First, the influence of prosody on segmental processing might not be immediate but hearing the disambiguating segmental materials after the onset overlap might have occurred before the analysis of the prosodic structure became effective. In that case, the failure to replicate the prosodic effect might be a function of the experimental task, which induces an early disambiguation. On the other hand, the effect of prosody observed in Experiment 2 might be tied to the repetitiveness of the 2AFC task, which could have influenced the listeners' performance. Moreover, the identification functions observed in Experiment 2 were never close to the floor or ceiling, and their shape was far from categorical, indicating that participants were relatively uncertain about their responses. In Experiment 4, we sought to explore those possibilities further by using a gating task in which the disambiguating segmental materials used in Experiment 3 were masked by noise. The absence of disambiguating cues allowed us to test whether the prosodic effect observed in Experiment 2 was a task-specific effect, or whether the null prosodic effect found in Experiment 3 was indeed a result of the disambiguating segmental materials coming into play earlier than the effect of prosody.

[8] In these cases, we can also be certain that these effect sizes are not inflated by selective reporting.

Experiment 4

In this experiment, participants performed a gating task, in which they heard part of a word and had to guess which word the speaker intended. The sentences used in Experiment 3 were presented with the later parts of the critical word replaced by a masking noise, and participants were asked to guess the intended word. The critical trials were those in which the participants heard only the earlier parts of words that overlapped phonologically. For instance, they heard a stimulus [ʔɑb] + MASKER and had to decide whether the intended word was *abjad* /ɑbjɑd/→[ʔɑbjɑt] or *qabel* /ʔɑbel/→[ʔɑbel]. Crucially, we tested whether this kind of task would again reveal a prosodic effect of preboundary lengthening, indicating that participants are more likely to assume that the intended word is a vowel-initial one when the preceding word carries cues to a prosodic boundary (i.e., the effect observed in Experiment 2 but not in Experiment 3).

To prevent participants from focusing exclusively on the context in the absence of disambiguating materials, additional fillers were used so that the segmental information would indeed allow participants to identify the intended word. For example, they heard both [ʔɑb] + MASKER and [ʔɑb**j**] + MASKER (with the additional segment [**j**]); the latter provided the bottom-up segmental support ([j]) needed to choose *abjad* over *qabel*. The inclusion of those trials was intended to prevent participants from adopting unusual strategies and instead focus on recognizing the intended word.

Method

Participants

16 students at the University of Malta participated in this experiment. They were native speakers of Maltese and Maltese English and participated for a small monetary compensation. There were 9 female and 7 male participants, aged 18 to 27.

Apparatus and stimuli

The apparatus was the same as in Experiment 3, but the eye-tracker was not used. The experimental stimuli from Experiment 3 were used. For each pair, two gating points were determined, one at which participants had little segmental information to distinguish the two members of a pair and one in which additional disambiguating information was present. For instance, for the pair *qabel* /ʔɑbel/ → [ʔɑbel] and *abjad* /ɑbɑd/→ [ʔɑbjɑt], the first splice point was at the release of the /b/, and the second one was 50 ms after that release (thus providing sufficient information about the following segment). It is possible that the first gate also provided some coarticulatory information about the following vowel ([ʔɑbjɑt] vs. [ʔɑbel]), but what is important is that the second gate (e.g., with [j] in [ʔɑbjɑt]) carries much clearer bottom-up phonetic support for [ʔ]*abjad* than for [ʔ]*abel*. The gated stimuli were followed by a complex masking sound with a base frequency of 70 Hz. Appendix A provides the locations of both gates for all stimuli used.

Procedure

All instructions were given on the computer screen as part of the experiment. Participants were told that they would hear sentences such as *Matthew jifhem* TARGET (Engl., 'Matthew understands TARGET'), in which the final word (the target) would be partly inaudible due to an additional sound. Participants were also told that they would see two words on the screen, and they would have to guess which of them was more likely to be the partially inaudible target word.

Each participant heard each of the 96 critical words (from 48 pairs) once each in of the two contexts (with and without preboundary lengthening). Additionally, they heard 48 filler trials, 24 of which used a glottal-stop initial target and 24 of which used a vowel-initial target.

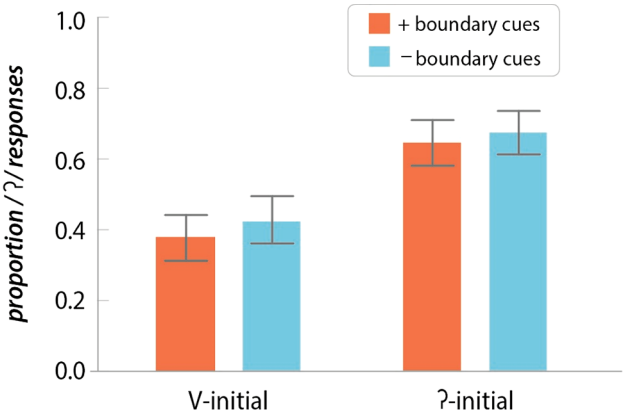

Fig. 7. Mean proportions of glottal-stop initial responses as a function of the prosody cue when the intended words were vowel-initial (left) and glottal stop-initial (right). The error bars are standard errors based on estimates from the *effects* package (Fox & Weisberg, 2018; Fox, 2003).

Results and discussion

We removed 57 trials (1.9%) from the analysis because of slow reactions (> 4s). In the filler trials with additional segmental information, participants guessed the intended word correctly 87.2% of the time. This indicates that participants engaged with the task, making use of available acoustic information, and focusing on identifying the intended words.

Fig. 7 shows the results from the experimental trials per condition. The figure displays the mean proportions of trials in which the glottal-stop initial word was chosen. The figure indicates that participants were to some extent able to use potentially residual fine phonetic cues in the gated stimuli because fragments stemming from glottal-stop initial words triggered more glottal-stop initial responses (around 60%) than vowel-initial word responses (around 40%). Crucially, the figure also shows a prosodic boundary influence: stimuli with preboundary lengthening gave rise to more vowel-initial responses (i.e., fewer /ʔ/-initial word responses) than stimuli without preboundary lengthening.

We tested this pattern for significance using a generalized mixed-effect model with a logistic linking function. The likelihood of a glottal-stop initial response was the dependent variable, and the prosody condition and stimuli source were independent variables. We used the independent variables as contrast coded fixed factors and also included their interaction. A random effect for participant and item was used with a full random-effect structure save for correlations between random effects. (Again, there is no random slope for the stimuli source over item because each item occurs on only one level of the stimulus-source factor.) Predictors were contrast coded with the prosody condition without preboundary lengthening and the glottal stop-initial stimulus source mapped on 0.5 and the prosody condition with preboundary lengthening condition and the vowel-initial stimulus source mapped on −0.5. Thus, the regression weights represent mean differences in the data, and the prediction is that the regression weights will be positive because both the no preboundary lengthening (minus-

Table 6
Results from the generalized linear mixed effect model for the likelihood of glottal-stop responses during Experiment 4.

	b (SE)	z	p
Intercept	0.128 (0.097)	1.320	0.187
Preboundary lengthening	0.173 (0.079)	2.184	0.029
Stimulus source	1.071 (0.175)	6.123	< .001
Preboundary lengthening : stimulus source	−0.082 (0.158)	−0.517	0.605

boundary) condition and the glottal stop-initial stimulus source are expected to give rise to more underlying glottal-stop responses. As shown in Table 6, the results indicate that both main effects of prosody (boundary) cue and stimulus source are significant in the expected direction, with no significant interaction between them.

The results therefore indicate that listeners are sensitive to the prosodic conditioning of glottal-stop insertion in Maltese. This is consistent with the main finding of Experiment 2, implying that the observed prosodic effect is not task-specific, but something that underlies speech perception. Apparently, when there is an indication of a prosodic boundary, participants perceived the glottal stop (or accompanying glottalization) as more likely to be a boundary marker than an underlying phoneme, reflecting the fact that speakers tend to insert an epenthetic glottal marking as a function of a prosodic boundary. Moreover, the fact that the significant prosodic effect arises in the absence of disambiguating segmental information after critical overlapping onsets is consistent with the view that the prosodic influence comes into effect relatively late in speech processing. The failure to observe the prosodic effect in Experiment 3 is thus attributable to the experimental task, in which the resolution of ambiguity occurred before the influence of prosody became effective.

General discussion

Our purpose in this study was to evaluate how listeners deal with phonetic variation in speech perception, focussing on two issues that have not yet received a great deal of attention: how listeners deal with variation at the onset of words in spoken-word recognition, and how the process might be prosodically conditioned. We identified the double function of the Maltese glottal stop as a phoneme and a glottal marker of prosodic boundaries as an interesting case to investigate these issues. First, we carried out a production experiment to answer two questions: Do epenthetic glottal stops occur frequently enough to pose a problem in spoken word recognition? Are there contextual or acoustic cues that might allow listeners to determine whether a given glottal stop is underlying or epenthetic? The answer to both those questions is yes: the glottal-stop epenthesis is frequent in Maltese, and there is a probabilistic prosodic contextual cue that distinguishes an epenthetic glottal stop from an underlying one. Epenthetic glottal stops are more likely to occur with preboundary lengthening that is consistent with a discernible prosodic boundary roughly corresponding to a minor phrase (roughly corresponding to a minor phrase larger than a phrase-internal word boundary, see Shattuck-Hufnagel & Turk, 1996) between the vowel-initial word carrying the epenthetic glottal stop and its preceding word. However, other possible cues, such as the phonological context (e.g., a V-V hiatus) and the phonetic detail (e.g., the duration and type of glottal gesture) do not appear to be used clearly to distinguish an epenthetic from an underlying glottal stop.

Experiment 2 used a 2AFC task and manipulated the presence of preboundary lengthening between a word that might have an epenthetic glottal stop and its preceding word. In line with the results from the production data, listeners were more likely to assume that the glottal stop was epenthetic if they heard a cue (preboundary lengthening) to a prosodic boundary. Experiment 3 used eye-tracking to evaluate whether that effect arises early during spoken-word recognition, but we found no early effect of prosody. We therefore considered two possibilities to account for the contradictory results between Experiments 2 and 3: the effect of boundary cues might occur at a late stage in processing and thus not have been captured in Experiment 3, or the observed prosodic effect in Experiment 2 might simply be a task-specific effect of the 2AFC task. Experiment 4 used the stimuli from Experiment 3 in a gating task with the disambiguating segmental information that followed the critical onset masked to observe whether prosody would have an effect later in spoken word recognition. The results of Experiment 4 indeed reveal that prosody has an effect, confirming that the influence of prosody is real but relevant only late in spoken word recognition.

Our results have implications for current theoretical issues in spoken-word recognition, especially with respect to two questions: how spoken words with phonological variants are recognized, and how spoken-word recognition is modulated by the interaction of prosodic and segmental processing. The first question has generally been discussed in terms of processing approaches versus representational approaches, in line with similar discussions in cognitive science, such as the question of whether visual object recognition relies on viewpoint-dependent or -independent representations (e.g., Tarr & Bülthoff, 1995). For instance, in line with an account that emphasizes representation, people might store different representations of an object to account for the rather different shapes that the object leaves on the retina depending on the angle of view (e.g., a road bike viewed from the front versus viewed from the side). Processing accounts, on the other hand, argue that the difference between the generically stored shape of an object and a given view is compensated for by processing that takes the viewpoint or orientation into account. Similarly, a phonological variant that arises when a vowel-initial word is produced with an epenthetic glottal stop could be recognized either by storing that phonological variant in the mental lexicon or by having processing mechanisms that filter out the epenthetic glottal stop before lexical access is attempted.

Our results show evidence for both types of processing. The results of Experiment 3 show that vowel-initial words are recognized relatively quickly, independent of prosodic boundary conditions. The acoustic mismatch between the surface form (i.e., a glottal stop-initial form) and the purported underlying form (i.e., a vowel-initial form) did not severely rule out the vowel-initial word as a lexical hypothesis, and listeners indeed perceived it as intended. Moreover, the recognition of the vowel-initial words with an epenthetic glottal stop took place despite the production data obtained in Experiment 1, which suggested that participants made no clear distinction in the phonetic detail of the glottal gesture used epenthetically to mark a vowel-initial word or as an underlying glottal stop. The results from both the production and perception experiments together support the view that a lexical representation of vowel-initial words includes a phonological variant with a glottal stop, so that the vowel-initial word can remain activated even if the variant is acoustically consistent with a competitor that has an underlying glottal stop.

A similar conclusion was reached by Mitterer and Reinisch (2015) for German. They found that German listeners are *slower* to recognize vowel-initial words when there is no glottal marking of the word onset. More important, this slow-down was similar to the slow-down that /h/-initial words suffered when they were produced without the initial /h/. Mitterer and Reinisch (2015) suggested that just as /h/, supposedly as a phoneme, needs to be represented lexically for a /h/-initial word, so does the glottal stop that occurs in vowel-initial words need to be represented lexically in German. Interestingly, Mitterer and Reinisch (2015) also compared the effect of deleting a glottal stop in German to that of deleting an underlying glottal stop in Maltese and reported that deleting a glottal stop in both languages led to similar reduction costs, which is again consistent with the view that vowel-initial words in German are represented with a glottal stop in the lexicon. Those results make us wonder whether spoken-word recognition by Maltese listeners would also be hindered by the absence of a glottal stop for vowel-initial words. Although that is an empirical question to be explored further by another study, comparing the production data between Maltese and

German could inform it. As noted in Mitterer and Reinisch (2015), a corpus of spontaneous speech in German indicates that about 90% of vowel-initial content words in German are produced with a glottal stop, comparable to the frequency of occurrence of glottal stops in underlying glottal stop-initial words in Maltese (about 97%), especially considering that the Maltese data come from a production task, whereas the German data are from a spontaneous speech corpus. Furthermore, only about 48% of vowel-initial words in Maltese show a glottal stop epenthesis, which strongly differentiates them from German vowel-initial words. The fact that German vowel-initial words are therefore more similar to Maltese glottal stop-initial words than to Maltese vowel-initial words suggests, as argued by Mitterer and Reinisch, that German uses the glottal stop as a phoneme, even though its distribution is position-specific. Thus, it appears that the phonological status of the glottal stop in German differs from that of a glottal stop used as a glottal marker of a vowel-initial word in Maltese. It is therefore reasonable to assume that deleting the glottal stop in vowel-initial words in Maltese might not pose reduction costs similar to those found in German.

As indicated by that discussion, the evidence supports the important role of storing pronunciation variants in the mental lexicon in recognizing vowel-initial words despite epenthetic glottal stops in Maltese. The results of the present study (especially those from Experiments 2 and 4) also inform the second question: how spoken word recognition might be influenced by an interaction between prosody and segmental processing. We found that listeners perceived a glottal stop (with the same acoustic input) as more likely to be a glottal marker of a vowel-initial word in the presence of a prosodic boundary cue (in the form of preboundary lengthening) than in its absence. Thus, in spoken-word recognition, listeners compute the prosodic structure (prosodic processing), possibly in parallel with segmental processing, and use the prosodic information in determining whether segmental information is driven lexically or post-lexically (prosodic-structurally). This possibility is indeed consistent with the view that segmental and prosodic processing of the speech signal are not independent and that speech perception is modulated by a computation of prosodic structure (Cho et al., 2007; Kim & Cho, 2013; Kim et al., 2018; Mitterer et al., 2016). Cho et al. (2007) suggested that listeners analyse available prosodic and segmental cues to compute the prosodic structure of the current utterance using the so-called *Prosody Analyzer*. They suggested that the detected prosodic boundaries inform the lexical competition process, so that although segmental processing determines the phonetic content of the current input (matching with the lexical hypothesis), the prosodic analysis indicates where words are likely to begin and end. Building on that idea, Kim et al. (2018) used an eye-tracking study to further demonstrate that a phonological variant that occurs due to a prosodically conditioned phonological process in Korean is recognized by listeners in reference to the prosodic structure being computed. They also provided some evidence that the resolution of potential ambiguity created by a phonological process comes relatively late in speech processing. Our results in this study are consistent with Kim et al.'s finding, implying that although the segmental and prosodic analyses may take place in parallel, their effects do not seem to come into play simultaneously: the segmental analysis activates all possible lexical hypotheses, and its activation is further modulated by the prosodic analysis at a relatively late stage in spoken-word recognition. This view is also in line with models that assume a relatively strong division in the relevant brain structures that process the segmental and prosodic aspects of speech (Giraud & Poeppel, 2012). Furthermore, this late effect of analysing the prosodic structure (as a higher-order linguistic structure) implies that lexical access takes place at multiple stages, comparable to the influence of syntactic information, which also comes into effect relatively late in spoken-word recognition (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). A similar late effect of context has recently been reported by Viebahn and Luce (2018), who found that non-canonical forms such as *winner* for *winter* are recognized better when presented in a casual-speech context. However, this effect arose only with slow reaction times, hence indicating that speech style influenced the lexical processing relatively late.

Interestingly, though, there is evidence that prosodic information can influence the lexical level relatively quickly. Salverda, Dahan, and McQueen (2003) showed that listeners are more likely to assume that the syllable *ham* embedded in a longer word *hamster* is the word *ham* (rather than the first syllable of the longer word) when the syllable is relatively long. This finding was evident in early fixations in an eye-tracking task, hence convincingly showing an early effect. This effect mirrors the fact that the syllable *ham* will receive word-final lengthening in *ham* but not in *hamster*, in accordance with temporal modification due to prosodic structure. While this effect could be explained by assuming that listeners recognize words better in their typical duration (following the idea that word-specific phonetic detail is stored in the mental lexicon, see Pierrehumbert, 2002), Shatzman and McQueen (2006) showed that this occurs even for newly learned words. This indicated that this effect is driven by prosodic analysis and not by listener's experience with how the actual phonetic realization of a given item may be influenced by prosodic structure in the language, because listeners had never heard this newly learned words in various prosodic contexts.

To summarize, our results show that both the storage of multiple variants in the mental lexicon and the processing of segmental and prosodic information contribute to spoken-word recognition, as evidenced by the recognition of glottal-stop-bearing variants of vowel-initial words in Maltese. The present study provides a theoretically interesting case in Maltese, a language that uses the glottal stop as both a phoneme and a glottal marker of vowel-initial words. This creates ambiguities that pose a possible problem in spoken-word recognition. Most crucially, our results imply that prosodic analysis comes into effect at a late stage in speech processing, modulating lexical competition in spoken-word recognition. We thus propose that theories of spoken-word recognition be refined to account for interactions between segmental and prosodic analyses and capture the role of interplay among phonetics, phonology, and prosody in speech perception.

Acknowledgements

We thank the students at the University of Malta who participated in the study. We also thank the editors Kathleen Rastle and Gareth Gaskell, and three anonymous reviewers for their insightful and constructive comments from which we have benefited greatly in improving the quality of our manuscript in various aspects. The work was supported by University of Malta Research Grant (CGSRP01-18) awarded to the first author, and by Global Research Network program through the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (Grant No. NRF-2016S1A2A2912410) awarded to the third author. The data and analysis scripts associated with four experiments can be found online at https://osf.io/pw74u.

Appendix A. Test words with the location of gates used in Experiments 3 and 4

See Table A1.

Table A1
Pseudo-onset overlap words used in Experiments 3 and 4.

Glottal-stop initial word	Vowel-initial word	First gate	Second gate
qabad	abaq	1/3 into vowel	closure
qabar	ghabra	/b/ release	50 ms after release
qabbad	abbati	/b/ release	/t/-release
qabel	abjad	/b/ release	50 ms after release
qabez	ghabbex	50 ms into closure	voice onset after /b/-release
qabru	abbuz	vowel midpoint	b release
qaddej	addocc	/d/-release	50 ms after release
qaddiefa	ghaddas	/d/-release	50 ms after release
qaddisa	addotta	/d/-release	50 ms after release
qadef	adda	50 ms into closure	voice onset after /d/-release
qadima	ghadira	1/3 into first vowel	1/3 consonant after second vowel
qafas	affari	40 ms after /f/ onset	20 ms after onset second vowel
qafla	afda	midpoint of /f/	after /d/ or /l/
qahba	ahbar	1/3 s into first vowel	offset of first vowel
qajjem	ajruport	1/3 of /j/	offset /j/ + 50 ms
qala	ghalqa	2/3 of first vowel	onset of second vowel
qalb	Alpi	/p/ release	/p/ release + 70 ms
qalziet	alkohol	2/3 /l/	after stop release
qamar	amment	40 ms after /m/ onset	20 ms into second vowel
qamh	alf	vowel midpoint	1/3 into /m/or /l/
qammiela	ammetta	2/3 of /m/	30 ms after second vowel onset
qanfud	anzjan	midpoint of /n/	30 ms after /n/ offset
qanpiena	annimal	midpoint of /n/	30 ms after /n/ offset
qarabaghli	arancina	midpoint of second vowel	30 ms after second vowel offset
qarben	gharbiel	/b/ release	plus 40 ms
qarn	art	midpoint of /r/	endpoint of /r/ + 30 ms
qarnit	armata	midpoint of /r/	endpoint of /r/ + 30 ms
qarrej	ardit	40 ms of /r/	endpoint of /r/ + 30 ms
qasab	assalt	40 ms of /s/	endpoint of /s/ + 30 ms
qasrija	ghasfur	midpoint of /s/	endpoint of /s/ + 30 ms
qassata	assistent	2/3 of /s/	endpoint of /s/ + 30 ms
qastan	astrat	before /t/-release	plus 70 ms
qatel	attent	40 ms into closure	voice onset of second vowel
qatra	atleta	before /t/-release	plus 90 ms
qattus	attur	1/3 of second vowel	40 ms after vowel offset
qawl	awtur	2/3 into first vowel	40 ms after vowel offset
qawsalla	Awstralja	midpoint of /s/	30 ms of second vowel
qawwi	Awissu	40 ms of /v/	30 ms of second vowel
qawwies	ghawwiem	midpoint of second vowel	30 ms of final consonant
qieghdin	editur	midpoint of second vowel	30 ms of final consonant
qishom	isqof	2/3 of /s/	30 ms of second consonant
qodma	ghodda	stop closure	stop release
qoffa	offra	2/3 of /f/	30 ms of voicing post /f/
qorti	ordni	midway stop closure	stop release
qoxra	ghoxrin	end of frication	plus 50 ms
quccija	ucuh	midpoint of frication	30 ms after onset of second vowel
quddiem	udjenza	after stop release	plus 90 ms
qurdiena	urgenti	midpoint of /r/	after stop release

Appendix B. Examples of speech tokens without glottalization and different versions of glottalization

See Figs. A1–A5.

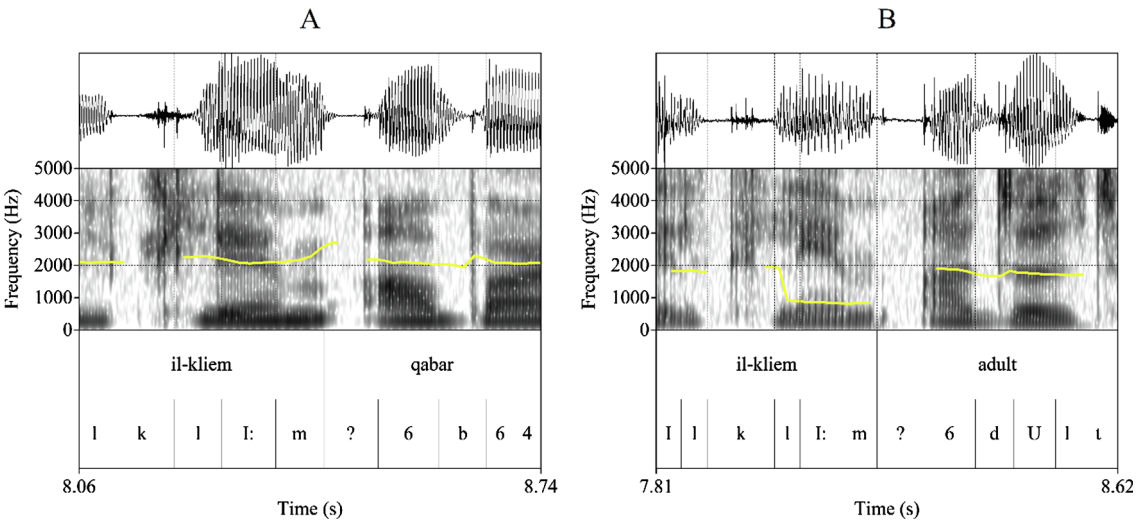

Fig. A1. Two examples of phrases in which there is a full glottal stop identified by the forced-alignment mechanism. Each panel shows the waveform (top row) and the spectrogram with an overlaid pitch curve (middle row) plus two transcriptions, an orthographic and a phonetic one using SAMPA. There is a clear (near-)silence at the word boundary, one triggered by the glottal stop-initial word *qabru* (Engl., 'tomb', **Panel A**) and one by a vowel-initial word *adult* (Engl., 'adult', **Panel B**).

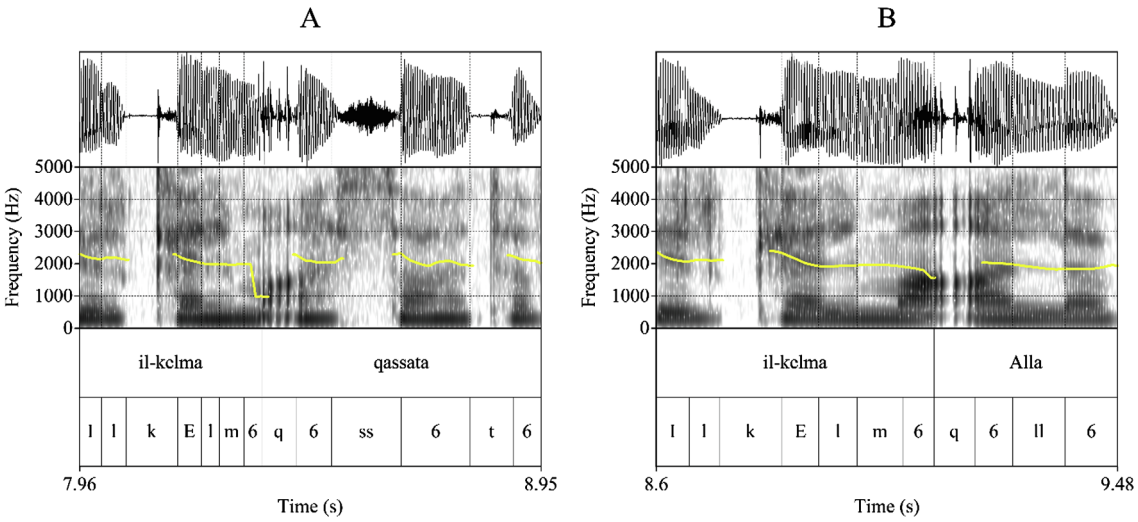

Fig. A2. Two examples of phrases in which there is a glottalization that is visible as a break in the pitch contour, but the forced-alignment did not find a full glottal stop. Duration has been estimated by raters (see main text). Each panel shows the waveform (top row) and the spectrogram with overlaid pitch curve (middle row) plus two transcriptions, an orthographic and a phonetic one using SAMPA. The pitch estimation algorithm fails to find a continuous pitch track at the word boundary, both for the glottal-stop initial word *qassata* (name of a typical Maltese pastry) in **panel A** and the vowel-initial word *Alla* (Engl., 'God' in **Panel B**. Note that the word *Alla* is not tied to a given religion and is used in Catholic services and hence relatively uncontentious).

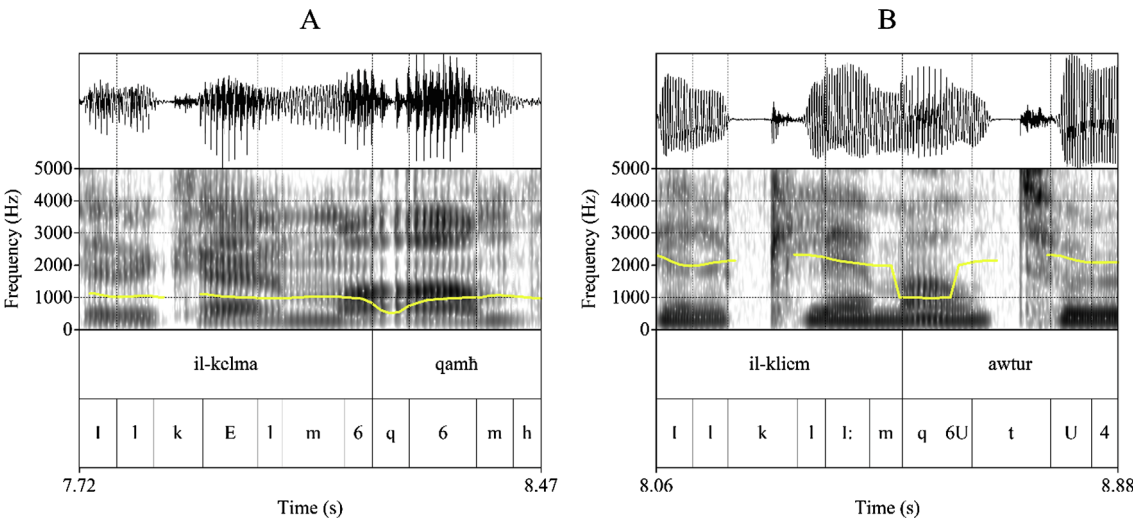

Fig. A3. Two examples of phrases with a sudden dip in the pitch curve, which was taken as evidence of glottalization despite a continuous f0 curve. Each panel shows the waveform (top row) and the spectrogram with overlaid pitch curve (middle row) plus two transcriptions, an orthographic and a phonetic one using SAMPA. The glottalization is visible in the sudden lowering of the pitch contour at the word boundary, which often goes hand in hand with a lowered amplitude (here visible in **Panel A**). As in the earlier figures, **panel A** provides an example from a glottal stop-initial word, *aamħ* (Engl., 'grain') and **Panel B** an example from a vowel-initial word, here *awtur* (Engl., 'author'). The duration of the glottalization has been estimated by raters (see main text).

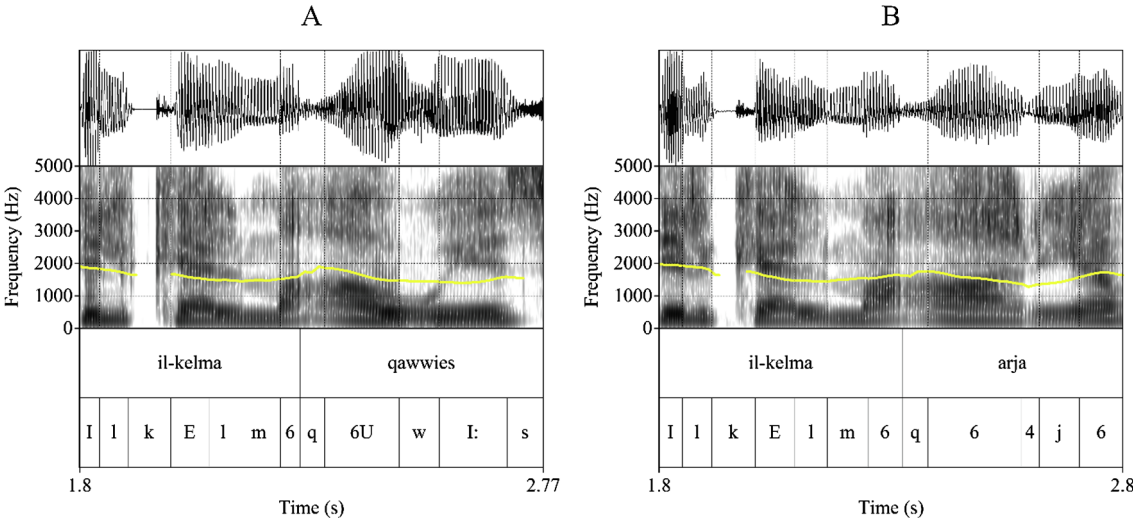

Fig. A4. Two examples of phrases with a sudden dip in the amplitude at the word boundary, which was taken as evidence of glottalization despite a continuous f0 curve. Each panel shows the waveform (top row) and the spectrogram with overlaid pitch curve (middle row) plus two transcriptions, an orthographic and a phonetic one using SAMPA. The glottalization is not visible in the pitch contour, which is relatively smooth. Instead, there is a sudden drop in the amplitude at the word boundary, which already had been reported as a means that some speakers use for glottalization. Interestingly, these examples are from the same speaker, and only two speakers from the eleven in the sample ever produced such a pattern. As in the previous figures, the example in **panel A** stems from a glottal stop-initial words, *qawwies* (Engl., 'archer') while the example in **Panel B** stems from the vowel-initial word *arja* (Engl., 'air').

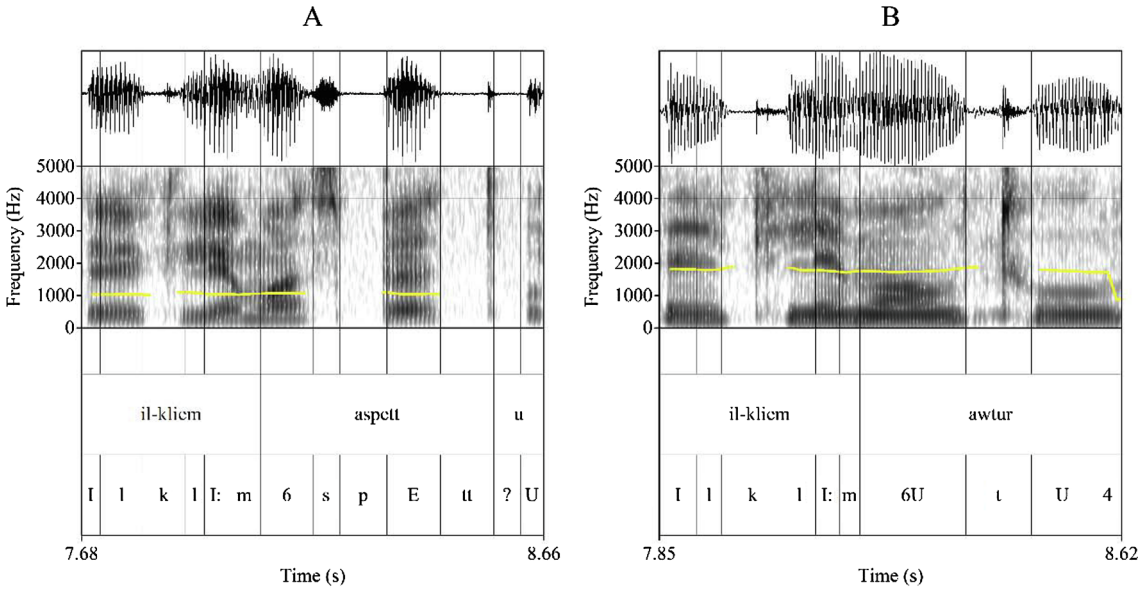

Fig. A5. Two examples of phrases in which the vowel-initial word is produced without any form of glottalization. Each panel shows the waveform (top row) and the spectrogram with overlaid pitch curve (middle row) plus two transcriptions, an orthographic and a phonetic one using SAMPA. There is no discontinuity of either the amplitude or the pitch at the word boundary before the vowel initial words *aspett* (Engl., 'aspect', **Panel A**) and *awtur* (Engl., 'author', **Panel B**). Panel A additionally shows the following word *u*, Enl. 'and' to document that even such an unaccented function word can sometime trigger glottal-stop epenthesis.

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2019.104034.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439. https://doi.org/10.1006/jmla.1997.2558.

Azzopardi-Alexander, M., & Borg, A. (1996). *Maltese* (1st ed.). London, New York: Routledge.

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

Booij, G. (1995). *The phonology of Dutch.* Oxford, UK: Clarendon Press.

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Can hearing puter activate pupil? Phonological competition and the processing of reduced spoken words in spontaneous conversations. *Quarterly Journal of Experimental Psychology, 65*(11), 2193–2220. https://doi.org/10.1080/17470218.2012.693109.

Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language, 62*(4), 421–437. https://doi.org/10.1016/j.jml.2010.01.002.

Bürki, A., & Gaskell, M. G. (2012). Lexical representation of schwa words: Two mackerels, but only one salami. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 38*(3), 617–631. https://doi.org/10.1037/a0026167.

Bybee, J. (2001). *Phonology and language use.* Cambridge: Cambridge University Press.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass, 10*(3), 120–141. https://doi.org/10.1111/lnc3.12178.

Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics, 64*, 71–89. https://doi.org/10.1016/j.wocn.2016.12.003.

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. https://doi.org/10.1016/j.wocn.2006.03.003.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language, 51*(4), 523–547. https://doi.org/10.1016/j.jml.2004.07.001.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review, 11*, 1084–1089. https://doi.org/10.3758/BF03196741.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* MIT Press.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*(4), 423–444. https://doi.org/10.1006/jpho.1996.0023.

Draxler, C., & Jänsch, K. (2004). SpeechRecorder – A universal platform independent multi-channel audio recording software. In Proceedings. of 4th intl. conference on language resources and evaluation.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*(4), 296–340. https://doi.org/10.1006/cogp.1999.0730.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software, 8*(15), 1–27. https://doi.org/10.18637/jss.v008.i15.

Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software, 87*(9), 1–27. https://doi.org/10.18637/jss.v087.i09.

Galea, L. (2016). *Syllable structure and gemination in Maltese. Ph.D dissertation.* Cologne, Germany: University of Cologne.

Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics, 45*, 106–113. https://doi.org/10.1016/j.wocn.2014.04.001.

Gaskell, M. G. (2003). Modeling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*, 447–463. https://doi.org/10.1016/S0095-4470(03)00012-3.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144–158. https://doi.org/10.1037//0096-1523.22.1.144.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 2*, 613–656. https://doi.org/10.1080/016909697386646.

Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1632–1647. https://doi.org/10.1037/a0011977.

Gatt, A., & Čéplö, S. (2013). Digital corpora and other electronic resources for Maltese. In A. Hardie, & R. Love (Eds.). *Corpus linguistics* (pp. 96–97). Lancaster: UCREL (Proceedings of Corpus Linguistics Conference, 96).

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517. https://doi.org/10.1038/nn.3063.

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

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133–159. https://doi.org/10.1006/jmla.2000.2764.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics, 65*, 575–590. https://doi.org/10.3758/BF03194584.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology, 15*(2), 189–226. https://doi.org/10.1017/S0952675798003571.

Kazanina, N., Bowers, J. S., & Idsardi, W. (2017). Phonemes: Lexical access and beyond.

Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-017-1362-0.
Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America, 134*(1), https://doi.org/10.1121/1.4807431 EL19–EL25.
Kim, S., Mitterer, H., & Cho, T. (2018). A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *PLoS ONE, 13*(8), e0202912. https://doi.org/10.1371/journal.pone.0202912.
Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language, 45*(Supplement C), 326–347. https://doi.org/10.1016/j.csl.2017.01.005.
Ladefoged, P., & Johnson, K. (2014). *A course in phonetics* (7th ed.). Stamford, CT: Cengage Learning.
Ladefoged, P., & Maddieson, I. (1996). *Sounds of the world's languages*. Oxford: Blackwell Publishers.
Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the lexicon. *Cognition, 38*, 245–294. https://doi.org/10.1016/0010-0277(91)90008-R.
Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
Maddieson, I., & Precoda, K. (1989). Updating UPSID. *The Journal of the Acoustical Society of America, 86*(S1), https://doi.org/10.1121/1.2027403 S19-S19.
Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 576–585. https://doi.org/10.1037/0096-1523.15.3.576.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0.
McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 539–553. https://doi.org/10.1037/0278-7393.29.4.539.
McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60*, 661–671. https://doi.org/10.1121/1.419865.
Mitterer, H. (2018). Not all geminates are created equal: Evidence from Maltese glottal consonants. *Journal of Phonetics, 66*, 28–44. https://doi.org/10.1016/j.wocn.2017.09.003.
Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics, 54*, 68–79. https://doi.org/10.1016/j.wocn.2015.09.002.
Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59*(8), 1395–1424. https://doi.org/10.1080/17470210500198726.
Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science, 30*(3), 451–479. https://doi.org/10.1207/s15516709cog0000_57.
Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73–103. https://doi.org/10.1016/j.wocn.2005.03.003.
Mitterer, H., Kim, S., & Cho, T. (2013). Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2013.02.001.
Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 244–263. https://doi.org/10.1037/a0012730.
Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception & Psychophysics.*. https://doi.org/10.3758/s13414-012-0407-8.
Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language, 85*, 116–134. https://doi.org/10.1016/j.jml.2015.08.005.
Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language, 98*, 77–92. https://doi.org/10.1016/j.jml.2017.09.005.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234. https://doi.org/10.1016/0010-0277(94)90043-4.
Ohala, J. J., & Ohala, M. (1995). Speech perception and lexical representation: The role of vowel nasalization in Hindi and English. In B. Cornell, & A. Arvanti (Eds.). *Phonology and phonetic evidence. Papers in laboratory phonology IV* (pp. 41–60). Cambridge: Cambridge University Press.
Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*, 2382–2393. https://doi.org/10.1121/1.2178720.
Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics, 69*, 1–11. https://doi.org/10.1016/j.wocn.2018.04.001.
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017.
Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.). *Laboratory phonology VII* (pp. 101–139). Berlin: Mouton de Gruyter.
Connine, C., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review, 23*(3), 235–245. https://doi.org/10.1515/TLR.2006.009.
Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61*, 19–36. https://doi.org/10.1016/j.jml.2009.02.005.
Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics, 29*(4), 407–429. https://doi.org/10.1006/jpho.2001.0145.
Roberts, A. C., Wetterlin, A., & Lahiri, A. (2013). Aligning mispronounced words to meaning: Evidence from ERP and reaction time studies. *The Mental Lexicon, 8*(2), 140–163. https://doi.org/10.1075/ml.8.2.02rob.
Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics, 43*, 11–25. https://doi.org/10.1016/j.wocn.2014.01.002.
Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51–89. https://doi.org/10.1016/S0010-0277(03)00139-2.
Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology, 3*, 371–405. https://doi.org/10.1017/S0952675700000695.
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247. https://doi.org/10.1007/BF01708572.
Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly-acquired words. *Psychological Science, 17*, 372–377. https://doi.org/10.1111/j.1467-9280.2006.01714.x.
Steffman, J. (2019). Intonational structure mediates speech rate normalization in the perceptoin of segmental categories. *Journal of Phonetics, 74*, 114–129. https://doi.org/10.1016/j.wocn.2019.03.002.
Steriade, D. (1999). Phonetics in phonology: The case of laryngeal neutralization. *UCLA working papers in linguistic: Vol. 2. Papers in phonology* (pp. 25–144). Los Angeles: UCLA.
Swinney, D. (1979). Lexical access during sentence comprehension: (Re)considerations of context effects. *Journal of Verbal Learning and Verbal Behaviour, 18*, 645–659. https://doi.org/10.1016/S0022-5371(79)90355-4.
Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 427–440. https://doi.org/10.1016/S0022-5371(79)90237-8.
Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1494–1505. https://doi.org/10.1037/0096-1523.21.6.1494.
Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001.
Viebahn, M. C., & Luce, P. A. (2018). Increased exposure and phonetic context help listeners recognize allophonic variants. *Attention, Perception, & Psychophysics, 80*(6), 1539–1558. https://doi.org/10.3758/s13414-018-1525-8.
Warner, N., & Weber, A. (2001). Perception of epenthetic stops. *Journal of Phonetics, 29*(1), 53–87. https://doi.org/10.1006/jpho.2001.0129.
Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020–2045. https://doi.org/10.1037/xge0000014.