Original Article

Language and Speech

Language *and* Speech
1–18

DOI: 10.1177/0023830920974401
journals.sagepub.com/home/las
SAGE

The Role of Segmental Information in Syntactic Processing Through the Syntax–Prosody Interface

Holger Mitterer
Department of Cognitive Science, University of Malta, Malta;
Hanyang Institute for Phonetics and Cognitive Science,
Department of English Language and Literature, Hanyang University, Korea

Sahyang Kim
Department of English Education, Hongik University, Korea

Taehong Cho
Hanyang Institute for Phonetics and Cognitive Science,
Department of English Language and Literature, Hanyang University, Korea

Abstract

In two experiments, it was investigated whether potentially contrastive segmental information in the form of an epenthetic glottal stop in Maltese can influence syntactic parsing decisions. The glottal stop in Maltese serves a dual function as a phoneme used for lexical contrast and a non-contrastive phone that may mark a prosodic juncture. In both experiments, participants perceived a larger prosodic boundary before the word *u* (Engl. "and") if the *u* was produced with an epenthetic glottal stop, showing the use of prosodically conditioned segmental information in syntactic parsing. Furthermore, listeners were generally unaware of the existence of the epenthetic glottal stop even though a glottal stop is used as a phoneme represented as a grapheme "q." They also perceived a larger prosodic juncture when the preceding syllable was lengthened before the word *u* ("and"). These findings were consistent regardless of whether the glottal stop reinforced a late-closure decision (Experiment 1) or an early-closure decision (Experiment 2). The results indicate that both segmental and suprasegmental information influences syntactic parsing decisions, demonstrating that the syntax–prosody interface is reflected along both the segmental and suprasegmental (duration) dimensions, which are mediated by the phonetics–prosody interface.

Keywords

Segmental effect, syntactic parsing, prosodic effect, syntax–prosody interface, phonetics–prosody interface

Corresponding author:
Taehong Cho, Hanyang Institute for Phonetics and Cognitive Science, Department of English Language and Literature, Hanyang University, 222 Wangshimni-ro, Seongdong-gu, Seoul, 133-791, Korea.
Email: tcho@hanyang.ac.kr

Introduction

In the tradition of the *Sound Pattern of English* (Chomsky & Halle, 1968), phonetics was simply considered to be governed by low-level biomechanical processes outside the realm of linguistics. A great deal of subsequent studies in the past several decades, however, have provided ample evidence that some of the non-contrastive phonetic phenomena are in fact under the speaker's control and systematically governed by language-specific phonetic "grammar" (Cho et al., 2017; Cho & Ladefoged, 1999; Keating, 1984, 1990; Kingston & Diehl, 1994). Additionally, phonetics may also interact with syntax (e.g., Cooper, 1976; Klatt, 1975; Michelas & D'Imperio, 2012). Studies on the interaction between phonetics and syntax have indeed shown that syntactic processing is likely to be influenced by how an utterance is phonetically realized, especially along the prosodic, or suprasegmental, dimension (Jun & Bishop, 2015; Steinhauer et al., 1999).

A generally accepted view available in the linguistic literature cited above is that the syntax–phonetics interaction is mediated by the syntax–prosody interface. For example, a particular syntactic structure may be reflected in the fine-grained phonetic detail of suprasegmental features (e.g., pitch, duration, amplitude) through the syntax–prosody interface—i.e., an intricate and systematic interaction between prosodic phrasing and syntactic parsing (Shattuck-Hufnagel & Turk, 1996). Many researchers (Hayes, 1989; Nespor & Vogel, 1986; Selkirk, 1984) indeed explored the interplay between prosodic and syntactic structures under the rubrics of Prosodic Phonology (or the Prosodic Hierarchy theory), and provided theoretical insights into how phonological units may be organized in a hierarchy of nested levels of prosodic constituency in relation to syntactic structure. Crucially, in the theory of Prosodic Phonology, a prosodic structure of an utterance is not construed as being isomorphemic to the syntactic structure. Instead, prosodic structure is assumed to be reparsed indirectly driven by syntactic structure, for example, in reference to the head–complement relation (Nespor & Vogel, 1986), the maximal projection (Selkirk, 1986), and the c-command relation (Hayes, 1989), although a prosodic structure above a prosodic word level may be parsed in relation to intonational phonology, independently of syntax (Jun, 1998). The evidence for such a prosodic structure, whether syntactically or intonationally driven, has often been provided in relation to domains of phonological sandhi rules which may apply across a word boundary. The prosodic-structurally bounded sandhi rules often induce segmental alternations. For example, Nespor and Vogel (1986) identified the Utterance (the largest prosodic unit assumed in Nespor & Vogel) as a prosodic domain of the flapping rule in American English; and Hayes (1989) suggested that the domain for the palatalization of /z/ before a post-alveolar sound (*Is Sheila*) is a Clitic Group (a prosodic constituent larger than the Prosodic Word but smaller than the Phonological Phrase, and it is blocked across the Clitic Group juncture). Thus, the theory of Prosodic Phonology has provided theoretical underpinnings for the postulation that the prosodic structurally conditioned variation of segmental realization informs syntactic structure in some perceptually meaningful ways.

A classic example of the relationship between segmental realization and syntactic structure may be found in the syntactic constraint on the *wanna* contraction—as in, *Who do you wanna dance with?* This segmental reduction process is blocked or disfavored when there is a trace for an underlying Wh-element that has been moved to the front, as in $\textbf{\textit{Who}}_i$ *do you want* $\textbf{\textit{t}}_i$ *to dance with Vlad?* (cf. Radford, 1997). A syntactic influence on phonetic implementation is also found for flapping (segmental reduction of /t,d/). An early study by Egido and Cooper (1980) suggested that flapping is mainly constrained by syntactic structure. However, a subsequent study by Parker and Walsh (1982) showed evidence that flapping is in fact constrained more systematically by intonational structure, often independently of syntactic structure (see de Jong, 1998; Wagner, 2010, for further evidence).

These studies cited thus far indicate that although there is a close relationship between syntax and prosody, prosodic structure is parsed in its own right (see also Beckman, 1996; Shattuck-Hufnagel & Turk, 1996, for reviews), and segmental variation that seems to be related to syntactic structure is indeed mediated by prosodic structure. Previous studies on the interaction of prosodic structure and syntactic structure in the sentence processing literature, however, have mainly focused on how syntactic structure may manifest itself in the phonetic realization of *suprasegmental* features and how the syntactically-conditioned suprasegmental variation may be used in sentence processing (e.g., Kentner & Féry, 2013; Petrone et al., 2017; Steinhauer et al., 1999; Watson & Gibson, 2005). In fact, another line of experimental phonetic research on the phonetics–prosody interface has demonstrated that prosodic structure is also signaled by gradient phonetic variation along the segmental dimension (e.g., Cho et al., 2017; Kim, Kim, et al., 2018), which is finer-grained than segmental alternations due to phonological sandhi rules that are applied within a prosodic constituent governed by prosodic hierarchy. This leads to a hypothesis that is to be tested in the present study: prosodic structurally conditioned segmental detail contains perceptual cues to syntactic structure through the syntax–prosody interface, which is in turn exploited by listeners in sentence processing.

Before we move on to elaborate on testing this hypothesis, it should be noted here that we define the terms "segmental" and "suprasegmental" in a specific way largely following Lehiste (1970). By "segmental" we refer to speech attributes of any discrete unit (as a phone) that can be identifiable physically in the stream of speech as being either consonantal or vocalic. By "suprasegmental," we refer to speech attributes of pitch, duration, and amplitude that may be superimposed on segments, extending over more than one segment (i.e., over syllables, words, or phrases). So, for example, as discussed in Lehiste (1970), if an acoustic feature such as duration is determined for a single segment, it is intrinsic to the segment, thus being segmental in nature (e.g., quantity contrast of phonologically long versus short vowels or a singleton versus a geminate). On the other hand, if the same acoustic duration feature is defined over a domain larger than a single phone (or a segment), it is suprasegmental in nature (e.g., prosodic boundary-related lengthening effects). Furthermore, we follow Lehiste (1970) and assume that an acoustic feature is suprasegmental when it is established in relation to items in sequence whereas it is segmental when it is defined without reference to the sequence of items. With this division line drawn between segmental and suprasegmental features, listeners may build prosodic and segmental structures independently (see also Cutler, 2012, for related discussion) but the listeners may still face the problem that no acoustic dimension may signal exclusively segmental or suprasegmental information (see McQueen & Dilley, in press, for a review). For example, F0 is typically considered to be an important correlate of prosodic structure, thus being suprasegmental, but in tone languages, it is used to make lexical distinctions, so that it serves a phonological function like a phonological segment does in non-tonal languages. Duration is also generally considered to be an important correlate of prosodic structure, thus being suprasegmental, although it may be specified for a single segment (phonologically long versus short vowels or singleton versus geminate consonants). But duration as a suprasegmental feature may also come into play in phonological and lexical processing. For example, given that a lexical boundary is aligned with a prosodic word boundary, which is partly expressed in duration (being suprasegmental extending over more than a single segment), duration also co-determines which words we are hearing (Salverda et al., 2003). Thus, perceiving the prosodic structure of a sentence in terms of its duration (Morrill et al., 2015) or pitch contour (Morrill et al., 2014) can influence listeners' spoken word recognition; and it may also heavily influence the decision on whether a segment is a singleton or a geminate (Mitterer, 2018b).

With the terms "segmental" and "suprasegmental" being defined as such, there is a wealth of speech production data showing that segmental information can signal prosodic structure, and that there is also a wealth of speech comprehension data showing that suprasegmental features associated with prosodic structure influence syntactic parsing. But there is hardly any study available in the literature which explores whether and how listeners may use segmental variation when making parsing decisions. One exception is the study by Scott and Cutler (1984). They demonstrated that American English listeners use flapping of a stop across a word boundary (e.g., *met Anne* /mɛt#æn/ → [mɛɾæn]) as a cue to the absence of a prosodic boundary. When confronted with ambiguous sentences such as *the day we met Anne was beautiful*, listeners are more likely to assume that *beautiful* refers to *day* (rather than *Anne*) if the /t/ in *met* is flapped.

An exploration of the relationship between segmental variation and syntactic parsing provides important theoretical implications for speech comprehension. Typically, syntactic parsing is assumed to be influenced by prosodic structural information that is manifested primarily along the suprasegmental dimension (see, e.g., Kjelgaard & Speer, 1999; Schafer et al., 1996; Schafer & Jun, 2002; Snedeker & Trueswell, 2003), which in turn is assumed to be analyzed independently of the segmental content of an utterance (Hickok & Poeppel, 2007; see also Cho et al., 2007). The evidence that detailed segmental information influences syntactic parsing would entail, for instance, the possibility that this information is retained in spoken-word recognition, in contrast to claims that listeners extract abstract phoneme information from the speech stream. For instance, in the example of Scott and Cutler (1984) provided above, the parser needs to have access to the phonetic form in which the /t/ is produced, which contrasts with proposals that listeners use abstract phonemes for lexical access (Bowers et al., 2016). Thus, testing the hypothesis (i.e., "segmental" information arising with the syntax–prosody interface being exploited by listeners in sentence processing) will inform these theoretical considerations.

We examine this possibility here using the glottal stop in Maltese as a testbed. According to the definition of the terms "segmental" and "suprasegmental" set forth for the purpose of the present study, the glottal stop is segmental in nature, but it is an unusual speech sound in many respects. First, it often surfaces without a full stop of the airflow through the vocal tract, but it may be phonetically realized as glottalization over an adjacent vowel. In fact, the latter is a common surface form unless the glottal stop is a geminate (Ladefoged & Maddieson, 1996; Mitterer, 2018a). Second, it occurs in most (if not all) languages, but with rather different linguistic functions, both within and between languages. In English, for example, it can be used as a segmental cue to coda voicing contrast (Penney et al., 2018) and also as a prosodic boundary marker that often accompanies vowel-initial words at larger prosodic boundaries (Redi & Shattuck-Hufnagel, 2001). The boundary-marking function is found in Italian as well, but with much lower frequency (Eger et al., 2019).

What additionally complicates the situation in Maltese is that a glottal stop not only is a phoneme by itself but can also mark otherwise vowel-initial words (Galea, 2016). This dual use of a glottal stop leads to potential ambiguity, since the Maltese listener has to decide whether the glottal stop is underlying or epenthetic (Mitterer et al., 2019). The nature of the glottal stop in Maltese differs from that of glottalizations observed in English (Garellek & Seyfarth, 2016). In English, glottalization (a form of creaky voice) may occur in association with coda /t/, which may be produced with glottal constriction either in the presence or absence of its oral constriction. Garellek and Seyfarth (2016) reported that the acoustic realization of glottalization is generally localized to the vicinity of the coda, which is segmental as it is defined for a particular segment. Garellek and Seyfarth reported another form of glottalization in English which may occur towards the end of a phrase (i.e., before a large prosodic boundary). The boundary-related glottalization was often found to extend over larger portions of phrases, indicating that it is suprasegmental.

Native listeners of English are indeed able to make phonetic distinctions between the two types of glottalization (Garellek, 2015) and use the glottalization phenomena straightforwardly in lexical processing (Heffner et al., 2013). This is, however, not the case for the Maltese glottal stop. With a database that exceeded those that found evidence for small residual phonetic cues (Roettger et al., 2014), a study by Mitterer et al. (2019) found no phonetic differences between epenthetic and lexical glottal stops, so that, for instance, both the epenthetic and lexical glottal stops were produced either with glottalization or with a full glottal stop with no difference between the two types. Listeners hence have no bottom-up information that allows them to decide whether a given glottal stop (realized with glottalization or with a full glottal stop) is signaling an underlying, lexical glottal stop or an epenthetic glottal stop. Importantly, Mitterer et al. (2019) further reported that listeners were not able to use the phonetic form of a glottal stop as a cue to whether the stop is underlying or epenthetic. That is, when a word-initial vowel triggered an epenthetic glottal stop, the words *attur* /ɑt:ur/ (Engl., "actor") and *qattus* /ʔɑt:us/ (Engl., "cat") effectively became cohort competitors so that listeners activated both words to the same extent when hearing [ʔɑt:u] (Mitterer et al., 2019).[1]

Mitterer et al. (2019) also showed that listeners might solve this problem in part by analyzing the prosodic structure of the utterance that bears on segmental processing. While speakers are more likely to use an epenthetic glottal stop if there is a larger prosodic boundary preceding the vowel-initial word, listeners assume that a glottal stop is more likely epenthetic than phonemic in the given prosodic context. This demonstrates that the glottal stop in Maltese serves the dual functions of providing segmental information for both lexical contrast and prosodic structuring. Here, we test whether the glottal stop, despite being a phoneme, can also influence sentence processing in Maltese.

In sum, the present study continues to explore possible interactions between phonetics and syntax by examining the effects of phonetic variation in Maltese (a Semitic language) on sentence processing, not only along the suprasegmental (prosodic) dimension that has been previously studied in other languages, but also along the segmental dimension that is to be newly explored here. We test the hypothesis that prosodic-structurally conditioned "segmental" information informs the syntax–prosody interface and is, therefore, exploited by listeners in sentence processing, which will illuminate the nature of the role of fine phonetic detail in sentence processing.

2 Experiment 1

We tested the potential of the glottal stop to influence parsing using ambiguous phrases of "coordinated names" (e.g., Kentner & Féry, 2013) such as *Malcom jew Gordon u Daniel* (Engl. "*Malcom or Gordon and Daniel*"). As (1) shows, there are two readings of this phrase.

(1) a. *[Malcom] jew [Gordon u Daniel]* "early closure": **[*N1*]** or [*N2* and *N3*]

b. *[Malcom jew Gordon] u [Daniel]* "late closure" : **[*N1* or *N2*]** and [*N3*]

Participants were asked to indicate which reading seemed more likely to them based on stimuli that differed in terms of two aspects: phrase-final lengthening (along the suprasegmental dimension) and the presence and absence of a glottal stop before a vowel-initial word (along the segmental dimension). The final syllable of a word (*Malcom* or *Gordon*) was lengthened or not lengthened before a potential syntactic juncture so that the final lengthening of *Malcom*, for example, was consistent with the "early-closure" parsing of (1a). The lengthening factor was also expected to reduce the demand characteristics of the experiment. Had we only manipulated the presence or absence of glottalization, listeners might use the glottalization cue just because there is nothing else in the experiment to respond to. Thus, with this additional manipulation, listeners were expected to meaningfully engage with the task without necessarily paying attention only to the glottalization cue. A glottal stop was inserted before the vowel-initial word /u/ ("and"), which is more consistent with the "late-closure" parsing of (1b) in which /u/ may form a new prosodic phrase. (Note that the Maltese word *jew* ("or") cannot be marked with a glottal stop, as it starts with a consonant /j/.)

Along with testing these possibilities, we also obtained further information from the listeners in order to examine their awareness of the cues they might use in making decisions. Listeners are usually unaware of non-contrastive epenthetic glottal stops, as was the case with German listeners (Mitterer & Reinisch, 2015). However, the presence of the grapheme /q/ in the language system may enhance awareness of the epenthetic glottal stop. Therefore, we asked participants after the experiment on what they based their decisions, to see whether some of them would mention the epenthetic glottal stop. This information is relevant for the discussion in the literature with respect to possible interactions between written- and spoken-language processing (Mitterer & Reinisch, 2015; Ziegler & Ferrand, 1998).

2.1 Method

Participants. Twenty-four native speakers (16 female) of Maltese participated in Experiment 1. They were students at the University of Malta, aged between 18 and 29.

Stimuli. A female native speaker of Maltese produced various versions of sentences such as *Malcom jew Daniel u Gordon* (*N1 or N2 and N3*). Using a Latin square, there were three possible orders of the three names in which each name appeared in each position (i.e., Gordon/Malcom/Daniel, Malcom/Daniel/Gordon, and Daniel/Gordon/Malcom). The speaker was asked to produce three different prosodic versions: without a prosodic break, with a break after the first name (*N1*), and with a break after the second name (*N2*). The first version (without a prosodic break) was used as a template for the experimental stimuli, and the latter two versions to gauge the typical amount of lengthening of syllables before a prosodic break, which was about 80% (see Table A1).

The tokens produced without a prosodic break were then annotated with the final syllables of *N1* and *N2* marked. Based on these annotations, experimental stimuli were generated using the PSOLA function in Praat (Boersma, 2001). Since the speaker did not produce glottalization for the vowel-initial word in these utterances (with no prosodic break inside the utterance), we added glottalization by manipulating pitch (F0), as difference in pitch is one of the primary phonetic parameters for glottalization accompanied by a glottal stop (e.g., Mitterer et al., 2019). To add glottalization to the conjunction /u/ (Engl. "and"), fundamental frequency (f0) was manipulated for the last 20ms of the preceding name (*N2* in the list) and the first half of the following *u*. For the last 20ms of *N2*, f0 was ramped down to 80Hz (with the speaker's normal pitch range of 150–250Hz), and then the following pitch points during /u/ were set as follows: 60Hz at 20%, 90Hz at 35%, and 80Hz at 50%, after which the original pitch was used again. Moreover, the amplitude in that part of /u/ was reduced to 50% (with a ramp of 15ms) as a lower amplitude is another phonetic property of glottalization (Hillenbrand & Houde, 1996). Given that these cues are typical for glottalization and not for a full glottal stop, it is important to remember that glottalization is a frequent implementation of both a lexical and an epenthetic glottal stop in Maltese which did not yield any discernible phonetic distinction (Mitterer et al., 2019). That is, the relative frequencies of glottalization and full glottal stops are roughly similar for epenthetic and lexical glottal stops.

For the duration along the suprasegmental dimension, the respective second syllables of the first two names (*N1* and *N2*) were lengthened from 0 to an increase of 80% in nine steps. For Step 1, (the second syllable of) *N1* was not lengthened and (the second syllable of) *N2* was lengthened by 80%; for Step 2, *N1* was lengthened by 10% and *N2* by 70%; for Step 9, *N1* was lengthened by 80% and *N2* was not lengthened at all. The final stimuli were hence longer than the original recording. The duration manipulation was done with PSOLA in Praat; "duration points" were introduced five milliseconds before and after the syllable onset and offset as marked in the transcriptions. The duration change was then ramped up and down over 10ms from one (i.e., no change) to the lengthening factor prescribed for the given syllable at the given step.

Note that by lengthening syllables in percentages according to their base duration introduces differences in terms of how much each syllable is lengthened in an absolute term. However, this proportional lengthening reflects Weber's law in that, for example, an increase in 40ms would be more noticeable on a 100ms base duration than a 160ms base duration. Figure 1 provides example stimuli of the first and last step of the duration continuum with and without the added glottalization. There were 54 different stimuli with three base sentences (three possible orders of the names), nine different duration levels, and two levels of glottalization (presence or absence).

Over the durational continuum, more "late-closure" decisions (i.e., [*N1* or *N2*] and [*N3*]) were expected for a lower step. For example, Step 1 with no final lengthening of *N1* but maximal final lengthening of *N2* was expected to induce a prosodic juncture after *N2* with more late-closure decisions. On the other hand, Step 9, which had the maximal final lengthening of *N1*, was more consistent with an "early-closure" parsing due to a prosodic juncture after *N1* (i.e., [*N1*] or [*N2* and *N3*]). The glottalization associated with /u/ ("and") was expected to induce a prosodic juncture after *N2* and, therefore, more late-closure decisions.

Procedure. The experiment was controlled by PsychoPy (Peirce et al., 2019) running on a standard PC. Participants were seated in front of a computer and provided with written instructions that explained how the phrase *Malcom jew Gordon u Daniel* can be understood in different ways. They were told that they would have to decide which of the two possible meanings was more likely given how the sentence was produced. The two alternatives were presented over three lines on the computer screen indicating the different possible groupings with either *N1* and *N2* on one line (*Malcom jew Daniel* ¶ *u* ¶ *Gordon*) or *N2* and *N3* on one line (*Malcom* ¶ *jew* ¶*Daniel u Gordon*).

Each of the 54 stimuli was presented three times. The stimuli were presented in three blocks consisting of three sub-blocks during which all 18 stimuli for a given basic order (e.g., *Malcom jew Daniel u Gordon*) were presented once in a randomized order. After finishing the experiment, participants were also asked questions about what differences they heard in the stimuli and what kind of information they thought they might have used in making their decisions.

2.2 Results and discussion

Since it sometimes took very long for participants to make a decision, we only used reaction times that fell within the interval of 0.2s to 6s. This led to the rejection of 168 cases, 4.32% of the data. Figure 2 shows how often participants made a "late-closure" decision (*[N1 or N2] and [N3]*) by grouping the first two names together. There was a small effect of the duration continuum on the decision, showing a tendency towards more late-closure decisions as the final lengthening of *N2* increased. This reflects that the final lengthening of *N2* is prosodically consistent with the syntactic parsing of [*N1* or *N2*] and [*N3*]. There was also an effect of glottalization across the board, showing around 15% more "late-closure" decisions when the /u/ ("and") after *N2* was produced with glottalization.

The data were statistically analyzed in R with a generalized linear mixed-effects model with a logistic linking function. Fixed effects were duration, glottalization (both contrast-coded), and

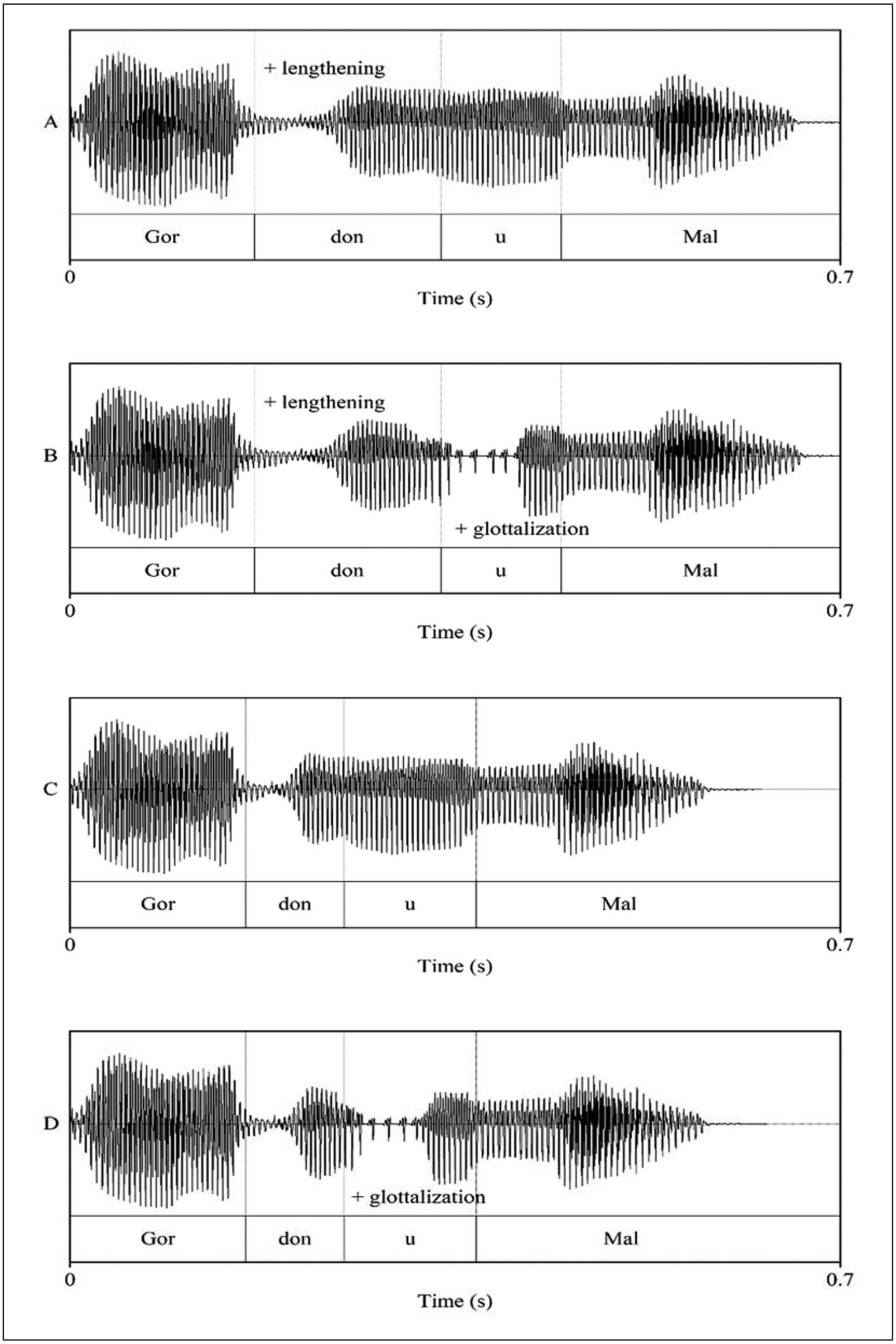

Figure 1. Examples of stimuli used in Experiment 1. The figures show the underlined part of *Daniel jew Gordon u Malcom* with the final syllable of *Gordon* (the *N2*) lengthened (Panel A and B) or not (Panel C and D) and glottalization added to the *u* (Panel B and D) or not (Panel A and C).

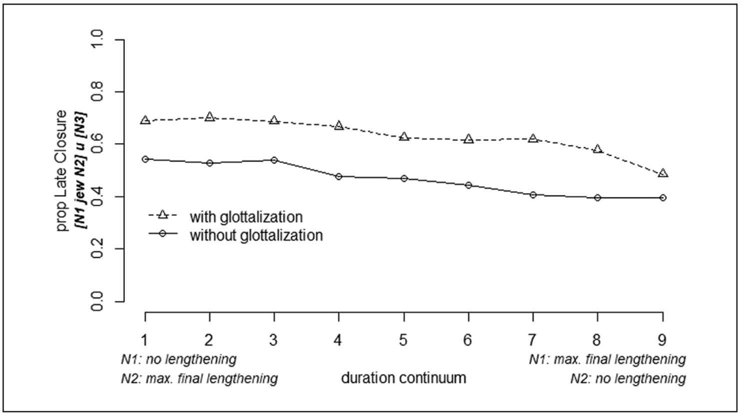

Figure 2. Proportion of late closure decisions (i.e., *[N1 jew N2] u [N3]*) in Experiment 1 over the duration continuum (Step 1 = no final lengthening of *N1* but maximal final lengthening of *N2*; Step 9 = maximal final lengthening of *N1* and no lengthening of *N2*) and on the glottalization of the *u* (Engl. "and").

their interaction. Participant was treated as a random effect with a full random effect structure, save for correlations between random effects (see Supplementary material for the exact formula). The results revealed a significant effect of duration ($b = -0.818$, SE $= 0.169$, $z = -4.843$, $p < 0.001$). The negative regression weight shows that there were fewer late-closure decisions with more final lengthening of *N1* (as the step became higher), indicating a prosodic juncture after *N1* (at the early-closure position). A significant positive effect of glottalization ($b = 0.811$, SE $=$ 0.260, $z = 3.121$, $p = 0.002$) indicates that participants made more late-closure decisions when the *u* was produced with glottalization, which helps mark prosodic juncture after *N2*, leading to a parsing of *[N1 or N2] and [N3].*

In experiments with multiple cues, a given cue tends to be more effective when the other cue is ambiguous (Harnad, 1987; Massaro, 1987). In the current case, this would mean that the glottalization cue is more efficient in the middle of the duration continuum. To capture this possibility, we added a predictor "duration squared" that also interacted with the glottalization predictor. Because this predictor is large at the margins and small in the middle of the continuum (due to centering of the duration predictor), the interaction of glottalization by duration squared would indicate such a pattern (cf. Mirman et al., 2008). In this additional model, however, this interaction was not significant ($b = -0.91$, SE(b) $= 0.78$, $z = -1.164$, $p = 0.244$), and adding this predictor did not lead to an improved model fit, $X^2(4) = 2.557$, $p = 0.643$.

When asked what they based their responses on, many participants indicated that they focused on how names were grouped together in line with the durational information, but not a single participant mentioned anything about the segmental variation due to the presence or absence of the epenthetic glottal stop. It appears that participants were not aware of the epenthetic glottal stop, despite the existence of the grapheme "q" in Maltese. This contradicts the assumption that an auditory input directly activates graphemes (Pattamadilok et al., 2014) and that learning to read makes listeners perceive speech in a more segmental manner (Ziegler & Ferrand, 1998). Instead, it seems that learning an orthography may influence meta-linguistic awareness but not the processing of the speech signal itself.

These results also indicate that the same glottal stop in Maltese which is realized along the segmental phonetic dimension and is used as a phoneme in the language can also be used to mark prosodic boundaries, which in turn influences syntactic parsing decisions. There is a long list of information sources that may influence syntactic parsing in the psycholinguistic literature (e.g., Steinhauer et al., 1999; Tanenhaus et al., 1995), and the development of the theory of Prosodic Phonology and Prosodic Hierarchy over the past several decades (Hayes, 1989; Jun, 1998; Nespor & Vogel, 1986; Selkirk, 1984) has provided theoretical insights into possible linguistic roles of segmental alternation in conjunction with the relationship between prosodic and syntactic structure. Nevertheless, as we discussed at the outset of the paper, the role of segmental detail has not yet been figured prominently in the sentence processing literature. Before we draw any firm conclusion, however, it is worthwhile to consider two issues with the current experiment. First, the glottalization reinforced the "late-closure" parsing, which is often considered to be a natural preference (the Late-Closure Principle, see Frazier & Rayner, 1982). Therefore, in Experiment 2, we changed the sentence by exchanging the locations of the conjunctions between *jew* ("or") and *u* ("and") so that the glottalization cue associated with *u* occurred earlier in the utterance (*[N1 and N2 or N3]*, e.g., *Daniel u Gordon jew Malcom*). In this coordinate construction, a prosodic juncture triggered by glottalization on the *u* would go against the Late-Closure Principle. Second, another weakness with Experiment 1 is that the effect of lengthening was found to be relatively small. Therefore, one cannot rule out the possibility that glottalization was relatively more audible, though not consciously perceived, and, therefore, it was the only reliable factor that participants could base their decisions on. Therefore, in Experiment 2, the duration manipulation was made more prominent to examine whether the glottalization cue still plays an independent important role even when the other (suprasegmental) cue is more salient.

3 Experiment 2

3.1 Method

Participants. Twenty-two native speakers of Maltese (13 female) participated in the experiment. They were aged between 19 and 31.

Stimuli and procedure. As in Experiment 1, we recorded the sentences from the same speaker in three versions: with no prosodic break, with an early break (i.e., after the first name, *N1*), and with a late break (i.e., after the second name, *N2*). The neutral versions with no break were produced with and without glottalization on the *u*, so we decided to use cross-splicing to generate the stimuli. Three utterances were used: two with glottalization on the *u* and one without it. One of the utterances with glottalization was used as a carrier, of which the *u* was replaced with the *u* from the other two utterances (one with and one without glottalization).

These utterances were further duration-manipulated using PSOLA as in Experiment 1 but with an increased range, so that at Step 1 the second syllable of *N1* (the first name) was shortened to 80% of its original duration and the second syllable of *N2* (the second name) was lengthened to 220% of its original duration. With each step, the compression factor for the first name was increased by 17.5% and reduced by the same amount for *N2* to bridge the range of 140% over nine steps. Consequently, at Step 9, the final syllable of *N1* was lengthened to 220% of its original duration and the final syllable of *N1* was shortened to 80% of its original duration. Figure 3 provides four examples of the stimuli with maximal bias for a late-closure parsing in Panel A (a short second syllable of *N1* and no glottalization) and a maximal bias for an early-closure parsing in Panel D (maximal lengthening of the second syllable of *N1* and glottalization on /u/). The procedure remained the same as in the previous experiment.

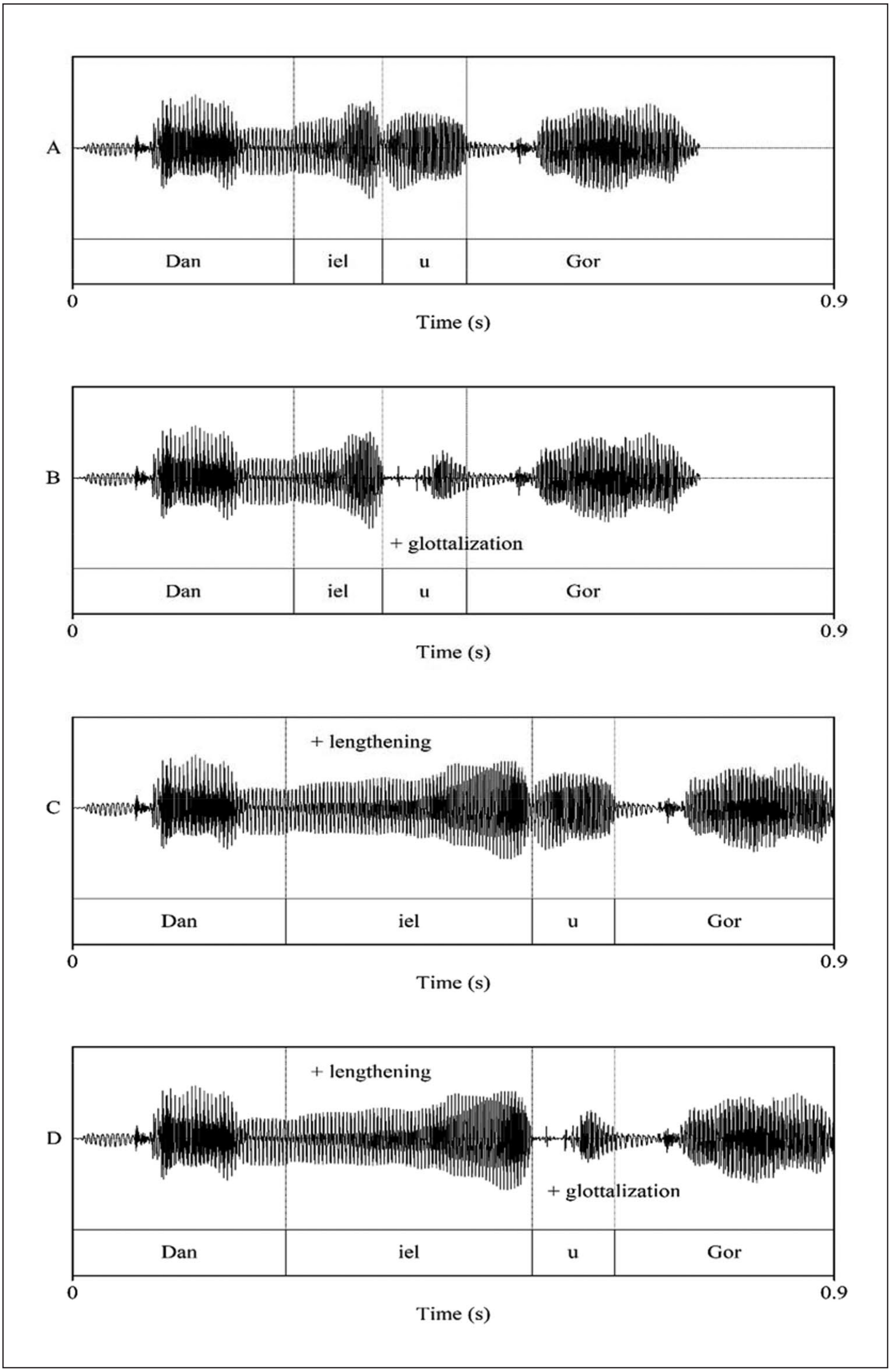

Figure 3. Examples of stimuli used in Experiment 2. The figures show the underlined part of *Daniel u Gordon jew Malcom* with the final syllable of *Daniel* (*N1*) lengthened (Panel C and D) or not lengthened (Panel A and B) and with glottalization for the *u* (Panel B and D) or not (Panel A and C).

3.2 Results and discussion

Similar to Experiment 1, none of the participants reported that they were aware of the presence of the epenthetic glottal stop as an influence on their decision. Instead, their answers (such as "two names were spoken together") suggested that they appeared to be aware of some kind of temporal relation.

Regarding the decisions, the data from one participant was rejected because s/he responded with the late-closure option for almost all the trials regardless of the conditions (on 160 of 162 trials). As in Experiment 1, we rejected trials for which the reaction time was outside the interval of 0.2s to 6s (187 cases, 5.39% of the data). The mean proportions of late-closure decisions for the remaining trials are provided in Figure 4. As was found in Experiment 1, the results show that the longer the final syllable of *N1* (i.e., along the duration continuum) the *lower* the likelihood of a late-closure response. The durational effect was generally more robust in Experiment 2 than in Experiment 1, which is consistent with the larger range of the durational manipulation in Experiment 2. Crucially, as for the effect of the "early" glottalization (*u* after *N1* rather than after *N2*), the results indicate that utterances with glottalization generally led to fewer late-closure responses. While this is numerically the opposite trend to Experiment 1 (as the locations of *u* ("and") and *jew* ("or") were switched), it shows the same effect of glottalization in such a way that participants tended to parse the sequence as [*N1*] *u* [*N2 jew N3*] when *u* ("and") had glottalization consistent with an epenthetic glottal stop. In other words, as in Experiment 1, the result indicates that listeners used the glottalization of *u* as a cue to a prosodic juncture, which in turn was likely used in syntactic parsing.

The same statistical analyses as in Experiment 1 were carried out using a generalized linear mixed-effects model with a logistic linking function and a maximal random-effect structure. The results confirmed that there were significant effects of duration ($b = 2.497$, SE $= 0.338$, $z = -7.376$, $p < 0.001$) and glottalization ($b = -0.378$, SE $= 0.142$, $z = -2.663$, $p = 0.008$) with no significant interaction ($b = 0.383$, SE $= 0.261$, $z = 1.470$, $p = 0.142$). As in Experiment 1, we also tested the predictor "duration squared" that also interacted with the glottalization predictor to test whether the effect of glottalization was larger at more ambiguous duration values. However, the interaction of glottalization by duration-squared was not significant ($b = -0.72$, SE(b) $= 0.89$, $z = -0.805$, $p = 0.421$), and adding this predictor did not lead to an improved model fit, $X^2(4) = 5.533$, $p = 0.237$.

The results in Experiment 2, therefore, confirm the early findings of Experiment 1, indicating that the segmental effect of glottalization on syntactic parsing is not an artifact. This means that the prosodic-structurally conditioned segmental variation influences syntactic parsing in a stable and systematic way even when the effects of other influencing factors may obscure the effect (e.g., the heightened phonetic clarity of durational information) or go in an opposition direction (e.g., glottalization supports an early-closure parsing while a late-closure parsing is generally preferred in the system).

4 General discussion

In two experiments, we demonstrated that listeners exploit potentially contrastive segmental information along the suprasegmental (prosodic) dimension. This was evident in how they parsed a structurally ambiguous coordinate structure: *N1 or N2 and N3* (*N1 jew N2 u N3* in Maltese) that could be parsed either with the first two names grouped together ("late-closure" parsing: *[N1 or N2] and [N3]*) or with the last two names grouped together ("early-closure" parsing: *[N1] or [N2 and N3]*). In Experiment 1, auditory stimuli were acoustically manipulated in such a way that *N1* or *N2* contained lengthening of the final syllable (in line with a phrase-final lengthening phenomenon along the suprasegmental dimension) and the conjunction *u* ("and") before *N2* was produced with glottalization, which may serve as a non-contrastive segmental marker of prosodic juncture.

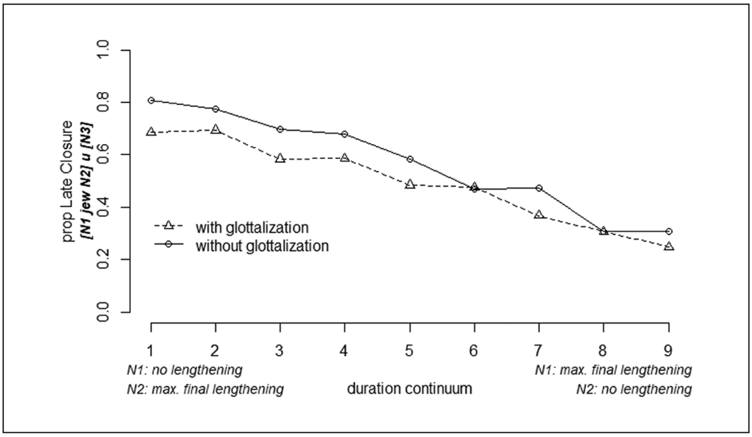

Figure 4. Proportion of late closure decisions (i.e., *[N1 u N2] jew [N3]*) in Experiment 1 over the duration continuum (Step 1 = no final lengthening of *N1* but maximal final lengthening of *N2*; Step 9 = maximal final lengthening of *N1* and no lengthening of *N2*) and on the glottalization of the *u* (Engl. "and").

Results of Experiment 1 showed that listeners utilized both the suprasegmental (durational) information and the segmental information of glottalization in making syntactic parsing decisions. Experiment 2 replicated this despite an enhancement of the durational difference along the suprasegmental dimension and the glottalization that led to an "early-closure" parsing running counter to the preferred late-closure parsing.

Our findings are in line with the view that the effect of fine phonetic detail on syntactic processing is not simply due to mapping of prosodic (suprasegmental) features on a particular syntactic structure but through the syntax–prosody interface. That is, the term prosody used in the "syntax–prosody mapping" or "syntax–prosody inference" does not simply refer to variation along the suprasegmental dimension (e.g., pitch, duration, and amplitude), but may be replaced by "prosodic structure" that refers to an abstract linguistic structure that organizes speech units in a hierarchical way (Beckman, 1996; Cho, 2016; Fletcher, 2010; Shattuck-Hufnagel & Turk, 1996). This means that syntax is not mapped directly on some suprasegmental features, but it may interact with prosodic structure. Since a prosodic structure is phonetically realized not only along the "prosodic" (suprasegmental) dimension but also the "segmental" dimension, the segmental effects observed in the present study support this structural view of the syntax–prosody interface.

Furthermore, listeners' answers to questions about what kind of information they may have used in making parsing decisions indicated that they were not aware that they used the segmental information of glottalization associated with an epenthetic glottal stop. This is intriguing because a glottal stop in Maltese is also used as a phoneme and represented as a grapheme "q." While this needs to be studied further in more rigorous ways than a post-experiment interview, it is worth noting that Maltese listeners are not fully aware of epenthetic glottal stops, in contrast with the assumption that experience with written language reshapes the pre-lexical processing of spoken language (e.g., Mitterer & Reinisch, 2015; Ziegler & Ferrand, 1998).

The findings of the present study raise an important question about how the syntactic parsing process can have access to this kind of segmental information along with prosodic and lexical information that were previously known to guide listeners' syntactic parsing. These results suggest that listeners do not discard allophonic segmental information. Recent accounts of spoken-word recognition indeed assume that the mental lexicon contains multiple word forms for a given word

(Bürki & Gaskell, 2012; Ranbom & Connine, 2007) and that pre-lexical processing does not abstract away from allophonic segmental variation that may be stored in the mental lexicon (Mitterer et al., 2013, 2018). The current result, therefore, resonates with such approaches, as is evident in the fact that segmental phonetic detail about pronunciation is available to guide listeners with syntactic parsing. More specifically, it appears that the Maltese mental lexicon contains the forms [u] and [ʔu] for the word *u* (Engl. "and") and which of these forms is activated influences the syntactic parsing decision one way or another.

There are, however, different ways in which the segmental information can guide parsing. Either the detailed segmental information is yet another type of information that is used in the constraint-satisfaction process of parsing or the segmental information influences a "Prosody Analyzer," as hypothesized by Cho, McQueen, and Cox (2007). These authors suggested that speech comprehension involves a simultaneous segmental and prosodic analysis of the speech signal. It is important to note that, even under this account, the glottalization cannot be analyzed immediately as carrying prosodic structural information because the epenthetic glottal stop could also be perceived as part of a glottal-stop initial word. The information, therefore, has to be fed into the segmental processing stream. Cho et al. (2007) further assumed that the so-called "Prosody Analyzer" computes the prosody structure, and the resulting information is mapped against the segmental information to accept or reject lexical hypotheses in competition. It is then possible that the "Prosody Analyzer" makes reference to the segmental information available in the speech signal that arises post-lexically, to compute a prosodic structure based on which the syntax–prosody mapping is determined. However, for both accounts it is necessary that the segmental processing does not discard allophonic variation so that this information can be used for either the syntactic parsing directly or influencing the prosodic-structural analysis.

It is also worthwhile considering that the analysis of prosodic structure influences the segmental analysis (Kim, Mitterer, et al., 2018; McQueen & Dilley, in press; Steffman, 2019) and that the segmental analysis influences the prosodic-structural analysis (Scott & Cutler, 1984). However, classical models of categorization cannot deal with such reciprocal influences (Smits, 2001), as there would be an infinite regress because no analysis ever is certain. McQueen and Dilley (in press) proposed a Bayesian model of prosody perception. It is an interesting question for future research how a Bayesian model of speech perception that not only includes the perception of words (Norris & McQueen, 2008) but also a prosodic analyzer would cope with the reciprocal influences.

The current results thus open up a whole new area of research on segmental influences on syntactic parsing, which will illuminate the nature of the intricate interactions between different components of linguistic structure. One specific question that may be immediately explored is how quickly segmental information driven by prosodic structure influences syntactic parsing relative to the segmental information used for lexical contrast. Answering such a question may also provide evidence for the use of segmental information at different levels of speech comprehension. This may be effectively tested by examining the impact of segmental information on the Event-Related Potential (ERP) component known as Closure-Positive shift (Steinhauer et al., 1999).

Acknowledgements

We thank the Maltese speakers and listeners for their participation in the experiments. We also thank the editor (Cynthia Clopper) and two anonymous reviewers for insightful and constructive comments and suggestions from which the study has greatly benefitted.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A2A03036736).

ORCID iDs

Holger Mitterer https://orcid.org/0000-0003-4318-0032
Taehong Cho https://orcid.org/0000-0002-8148-745X

Supplemental material

All stimuli, data, and analyses scripts are available at: https://osf.io/jeqdg/

Note

1. One might wonder whether the word *attur* might in fact also have an underlying glottal stop and be only vowel-initial in its orthographical form. However, *attur* (just as other vowel-initial words) triggers phonological processes that require the assumption that it is vowel-initial. The definite article in Maltese *(i)l* contains an initial vowel that is only used for consonant-initial words so that we get *il-qattus* (Engl., "the cat") but *l-attur* (Engl., "the actor").

References

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*(1–2), 17–68.

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

Bowers, J. S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, *87*, 71–83. https://doi.org/10.1016/j.jml.2015.11.002

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

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., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

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

Chomsky, N., & Halle, M. (1968). *The sound pattern of english*. Harper and Row.

Cooper, W. E. (1976). Syntactic control of timing in speech production: A study of complement clauses. *Journal of Phonetics*, *4*(2), 151–171. https://doi.org/10.1016/S0095-4470(19)31235-5

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

de Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, *26*(3), 283–310. https://doi.org/10.1006/jpho.1998.0077

Eger, N. A., Mitterer, H., & Reinisch, E. (2019). Learning a new sound pair in a second language: Italian learners and German glottal consonants. *Journal of Phonetics*, *77*, 100917. https://doi.org/10.1016/j.wocn.2019.100917

Egido, C., & Cooper, W. E. (1980). Blocking of alveolar flapping in speech production: The role of syntactic boundaries and deletion sites. *Journal of Phonetics*, *8*(2), 175–184. https://doi.org/10.1016/S0095-4470(19)31462-7

Fletcher, J. (2010). The prosody of speech: timing and rhythm. In W. J. Hardcastle, J. Laver & F. E. Gibbon (Eds.), *The handbook of phonetic sciences, second edition* (pp. 523–602). Blackwell Publishers.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

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

Garellek, M. (2015). Perception of glottalization and phrase-final creak. *The Journal of the Acoustical Society of America*, *137*(2), 822–831. https://doi.org/10.1121/1.4906155

Garellek, M., & Seyfarth, S. (2016). Acoustic differences between English /t/ glottalization and phrasal creak. *INTERSPEECH*, 1054–1058.

Harnad, S. (1987). *Categorical perception: The groundwork of cognition*. Cambridge University Press.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Rhythm and meter* (pp. 201–260). Academic Press. https://doi.org/10.1016/B978-0-12-409340-9.50013-9.

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, *28*(9), 1275–1302. https://doi.org/10.1080/01690965.2012.672229

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hillenbrand, J. M., & Houde. (1996). Role of F0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1182–1190. https://doi.org/10.1044/jshr.3906.1182

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

Jun, S.-A., & Bishop, J. (2015). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (pp. 217–240). Springer International Publishing. https://doi.org/10.1007/978-3-319-12961-7_12.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319. https://doi.org/10.2307/413642

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics*, *18*(3), 321–334. https://doi.org/10.1016/S0095-4470(19)30377-8

Kentner, G., & Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review*, *30*(2), 277–311. https://doi.org/10.1515/tlr-2013-0009

Kim, S., Kim, J., & Cho, T. (2018). Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics*, *71*, 65–80. https://doi.org/10.1016/j.wocn.2018.07.004

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

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454. https://doi.org/10.2307/416481

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194. https://doi.org/10.1006/jmla.1998.2620

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, *3*(3), 129–140. https://doi.org/10.1016/S0095-4470(19)31360-9

Ladefoged, P., & Maddieson, I. (1996). *Sounds of the world's languages*. Blackwell Publishers.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Massaro, D. W. (1987). Categorical partition: A fuzzy-logic model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–283). Cambridge University Press.

McQueen, J. M., & Dilley, L. (in press). Prosody and spoken-word recognition. In C. Gussenhoven & A. Chen (Eds.), *The handbook of prosody*. Oxford University Press.

Michelas, A., & D'Imperio, M. (2012). When syntax meets prosody: Tonal and duration variability in French accentual phrases. *Journal of Phonetics*, *40*(6), 816–829. https://doi.org/10.1016/j.wocn.2012.08.004

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

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

Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language*, *108*, 104034. https://doi.org/10.1016/j.jml.2019.104034

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*(Supplement C), 77–92. https://doi.org/10.1016/j.jml.2017.09.005

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*, 356–361. https://doi.org/10.1016/j.cognition.2013.07.011

Morrill, T. H., Baese-Berk, M., Heffner, C., & Dilley, L. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review*, *22*(5), 1451–1457. https://doi.org/10.3758/s13423-015-0820-9

Morrill, T. H., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition*, *131*(1), 69–74. https://doi.org/10.1016/j.cognition.2013.12.006

Nespor, M., & Vogel, I. (1986). Prosodic phonology. In *Prosodic Phonology*. De Gruyter Mouton. https://www.degruyter.com/view/title/31451

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Parker, F., & Walsh, T. (1982). Blocking alveolar flapping: A linguistic analysis. *Journal of Phonetics*, *10*(3), 301–314. https://doi.org/10.1016/S0095-4470(19)30991-X

Pattamadilok, C., Morais, J., Colin, C., & Kolinsky, R. (2014). Unattentive speech processing is influenced by orthographic knowledge: Evidence from mismatch negativity. *Brain and Language*, *137*, 103–111. https://doi.org/10.1016/j.bandl.2014.08.005

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-01193-y

Penney, J., Cox, F., Miles, K., & Palethorpe, S. (2018). Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics*, *66*, 161–184. https://doi.org/10.1016/j.wocn.2017.10.001

Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, *61*, 71–92. https://doi.org/10.1016/j.wocn.2017.01.002

Radford, A. (1997). *Syntactic theory and the structure of English: A minimalist approach*. Cambridge University Press.

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298.

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

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

Schafer, A. J., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, *11*(1–2), 135–163. https://doi.org/10.1080/016909696387240

Schafer, A. J., & Jun, S.-A. (2002). Effects of accentual phrasing on adjective interpretation in Korean. In M. Nakayama (Ed.), *East Asian language processing* (pp. 223–255). Center for the Study of Language and Information.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 450–466. https://doi.org/10.1016/S0022-5371(84)90291-3

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press.

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

Smits, R. (2001). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics*, *63*(7), 1109–1139.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130. https://doi.org/10.1016/S0749-596X(02)00519-3

Steffman, J. (2019). Intonational structure mediates speech rate normalization in the perception of segmental categories. *Journal of Phonetics*, *74*, 114–129. https://doi.org/10.1016/j.wocn.2019.03.002

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191–196. https://doi.org/10.1038/5757

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. https://doi.org/10.1126/science.7777863

Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, *28*(1), 183–237. https://doi.org/10.1007/s11049-009-9086-0

Watson, D., & Gibson, E. (2005). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, *59*(2–3), 279–300. https://doi.org/10.1111/j.1467-9582.2005.00130.x

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, *5*, 683–689. https://doi.org/10.3758/BF03208845

Appendix 1

Table A1. Properties of the nouns in the material set recorded for Experiment 1. The cells provide the duration of the first and second syllable of a noun in milliseconds plus the proportion of cases in which the noun was followed by a pause.

	Early boundary syl1 + syl2 / pauses	No boundary syl1 + syl2 / pauses	Late boundary syl1 + syl2 / pauses
N1	235+238 / 0.50	210+151 / 0.00	200+137 / 0.00
N2	208+138 / 0.00	210+150 / 0.00	226+230 / 0.57

Table A2. Properties of the nouns in the material set recorded for Experiment 2. The cells provide the duration of the first and second syllable of a noun plus the proportion of cases in which the noun was followed by a pause.

	Early boundary syl1 + syl2 / pauses	No boundary syl1 + syl2 / pauses	Late boundary syl1 + syl2 / pauses
N1	238+228 / 0.11	239+149 / 0.00	241+138 / 0.00
N2	240+239 / 0.00	249+211 / 0.00	263+291 / 0.33