Open Library of Humanities

Limits of audience design: Epenthetic glottal stops in Maltese

Clive Sciberras, Department of Cognitive Science, University of Malta, MT, clive.sciberras.14@um.edu.mt

Holger Mitterer, Department of Cognitive Science, University of Malta, MT; Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS), MT, holger.mitterer@um.edu.mt

Many languages tend to mark vowel-initial words with a glottal stop in connected speech, especially when that word is in a prominent position (Garellek, 2014). This also happens in Maltese, even though the glottal stop here also occurs as a phoneme, so that the epenthetic glottal stop may significantly alter which other words the vowel-initial word is similar to. For the pair *attur* /ɑt:ur/ versus *qattus* /ʔɑt:us/ (English 'actor' versus 'cat'), adding an epenthetic glottal stop to the vowel-initial word renders the two words more similar. This provides an interesting test bed for the hypothesis of audience design. If speakers want to highlight the contrast between the two words, they should, in contrast to what usually happens when a word is prominent, not produce an epenthetic glottal stop. We tested this in a production experiment and found that speakers instead produce much more glottal stops for such vowel-initial words under a phonological contrast compared with a given version, and even slightly more than when under a lexical contrast (e.g., *actor* versus *theatre*). Our results provide an example of a limitation of audience design.

1. Introduction

Every utterance is aiming to achieve two things: First, the production should be timely, adequate, and fluent. Secondly, an utterance needs to be understood (for a similar framing, see Ferreira & Dell, 2000). One of the continuing debates in psycholinguistics is to what extent speakers can incorporate a model of the listener into their production planning and shape their message to facilitate comprehension (for a recent review, see Jaeger & Buz, 2018). In this paper, we investigate to what extent this is the case on the phonetic level, using a case in which an apparent enhancement may in fact be counterproductive.

Clearly, speakers can modify their speech output if asked to speak clearly (for a review, see Smiljanić & Bradlow, 2009). Speakers then make a couple of adaptations (slower speech rate, expansion of the vowel space, etc.) which make it easier for both hearing-impaired and normal-hearing listeners to understand that type of speech.

The question to what extent speech production is catering for the listener plays, however, on multiple levels. To provide some examples, Ferreira and Dell (2000) investigated whether speakers would mention an optional word more often if the sentence otherwise was—temporarily—ambiguous (e.g., as in *I know (that) you missed practice*), but found no such effect. Similarly, Kraljic and Brennan (2005) investigated whether prosodic breaks are more likely in case of a potential ambiguity. Both studies found no effect of ambiguity on speakers' choices and suggested that speakers do not routinely take the listener into account when shaping utterances.

However, other studies have been more positive about speakers' ability to take the listener into account. Lockridge and Brennan (2002) used a story-telling task, in which actions were performed with typical or atypical objects (e.g., a knife versus an ice-pick for stabbing someone) and found that atypical objects were more likely to be explicitly mentioned than typical objects. However, this finding by itself would not have been sufficient to suggest that speakers design their utterance specifically for their audience. The mention might simply have been the consequence of taking longer to encode the unusual objects, which would have made them more prominent during conceptualization. This example nicely illustrates that apparently speaker-oriented choices may not be driven by the consideration of the listener after all. To convincingly show that these mentions of atypical objects were indeed driven by audience-design, Lockridge and Brennan (2002) manipulated whether listeners and speakers could both see a cartoon depicting the objects. The higher likelihood of atypical objects to be mentioned was strongly elevated when only the speakers but not the listeners could not see the cartoons, suggesting that the utterances were indeed specifically designed with the listeners' knowledge in mind.

One recurring issue in this debate remains whether apparent listener-oriented processes are in fact truly listener-oriented or can be explained by other variables (e.g., speakers' planning processes). This is especially so in the realm of phonetics and phonology. It has long been

recognized that the tendency for consonants to undergo place assimilation in production is inversely related to how salient place information is in perception (Hura, Lindblom, & Diehl, 1992; Steriade, 2001). This minimizes the impact of the articulatory simplification that assimilation has on the listener. As a consequence, nasal consonants are more likely to assimilate than stops; within the class of stops, languages with released stops are less likely to assimilate those stops than languages with unreleased stops, and fricatives are least likely to be assimilated, with the exception of the confusable /s/-/ʃ/ pair (Hura et al., 1992; Steriade, 2001). This is in line with confusion matrices, which indicate, that in final position, fricatives are less confusable than stops, which in turn are less confusable than fricative (see, e.g., Table II in Cutler, Weber, Smits, & Cooper, 2004). Similarly, vowel spaces seem to be optimized for perception, so that the vast majority of five-vowel inventories of the languages of the world look quite similar, targeting a maximal spread in the vowel space (Schwartz, Boë, Valleé, & Abry, 1997). However, it has been argued that such patterns may simply be a result of a cultural evolution of the languages themselves, rather than an (online) adaptation of speakers to listeners (de Boer, 2000).

Research has also indicated that words tend to be acoustically reduced when they are salient in the preceding discourse context or overall high in lexical frequency (Aylett & Turk, 2016; Ernestus, 2014). This has been found in corpora (Turnbull, 2018) as well as in experimental production studies (e.g., Burdin, Turnbull, & Clopper, 2014). Again, this makes sense for listeners as well, as they do not need a lot of bottom-up support if the context strongly suggests that this word is coming up. Vice versa, words that are not primed well by the context will be produced with full phonetic detail. Turnbull, Seyfarth, Hume, and Jaeger (2018) found that place assimilation in fluent speech (e.g., *lean bacon* produced as *leam bacon*) may also be driven by communicative purposes. They used the Buckeye corpus of spontaneous speech (Pitt et al., 2007) and measured the likelihood and strength of assimilation and related this to the contextual predictability of both the potential trigger word (i.e., *bacon*) and the potentially assimilated word (i.e., *lean*) and found that unpredictable trigger words were likely to trigger assimilation, while unpredictable target words were likely to block assimilation. This is useful for the listener because unpredictable target words that need bottom-up support receive it. On the other hand, unpredictable trigger words gain additional cues if they trigger assimilation, since listeners can treat the assimilation as a cue for the upcoming segment (Gow, 2003). Turnbull et al. (2018) argue that these results are compatible with a communicative account. However, they are also compatible with the assumption that contextual predictability influences retrieval effort for speakers, which is then translated into gestural strength during articulation.

Additional research, however, provides clearer evidence for intended audience design. Schertz (2013) asked participants to interact with a (fictitious) automatic speech recognizer, which asked feedback questions which were either specific ("Did you say *bit*?") or unspecific ("What did you say?"). Specific questions led speakers to manipulate specific acoustic cues, such

as Voice Onset Time (VOT) to highlight the contrast with the target *pit*. Seyfarth, Buz, and Jaeger (2016) used a similar methodology by presenting speakers with three words and then asking them to produce one of them. The critical manipulation was whether the other words visible on the screen included a particular competitor (e.g., the target *doze* accompanied or not by the competitor *dose*). Seyfarth et al. found that the presence of the phonologically similar competitor led to the enhancement of cues to final voicing. Buz, Tanenhaus, and Jaeger (2016) found that such effects can further be enlarged by feedback questions. All these studies show evidence for dynamic adaptation that indicates listener-oriented production.

However, the evidence indicates that not all phonetic parameters may be open to dynamic hyper-articulation. Schertz (2013) found that vowel quality was not specifically enhanced when the discrepancy between the target word and the apparently understood word involved a vowel difference (e.g., *bit* versus *bet*). Interestingly, this cannot be ascribed to an inability on the side of the speaker to modify formant frequencies, as they do so in altered-auditory feedback experiments (e.g., Katseff, Houde, & Johnson, 2012) as well as vowel spaces are larger in clear speech (Smiljanić & Bradlow, 2009).

There is also another example in which the enhancement that speakers provide may in fact be counterproductive. Cho, Kim, and Kim (2017) investigated how coarticulatory nasalization was influenced by prosodic factors. This included making a semantic or a phonological contrast leading to more prominence on the target word (e.g., *ban*). In a semantic contrast, a given word contrasts with an antecedent in terms of meaning (e.g., the word *ban* in the dialogue "Did Bob say access?" → "No, Bob said ban?"), while a phonological contrast requires a similarity in form (e.g., "Did Bob say bad?" → "No, Bob said ban?"). Unsurprisingly, Cho et al. found that under both contrasts, participants produced the target word *ban* with a longer duration and a higher pitch. Moreover, they found that they also spoke more clearly in the sense that the amount of nasal coarticulation was reduced on the /æ/ in *ban*. That is, the vowel in *ban* was less nasalized under contrast. While producing each segment clearer and with less influence of surrounding segments may generally benefit comprehension, in the case of *bad* and *ban*, it actually means that the words become more similar, because the contextually driven vowel nasalization makes them in fact more distinct. Given that Cho et al. (2017) found similar effects for phonological and lexical contrast, this provides an example where a typical enhancement is in fact counterproductive; yet, speakers still engage in that counterproductive enhancement.

Such findings provide evidence against the idea that speakers can fully model the listener when shaping their message. In this paper, we make use of another phenomenon where a default enhancement strategy has counterproductive consequences for the listener and ask the question whether speakers will nevertheless use such a counterproductive enhancement. This

becomes possible in Maltese because the glottal stop (ʔ) occurs in at least two functions. First of all, in Maltese, the glottal stop is a phoneme and can occur in onset and coda position in a syllable (Azzopardi-Alexander & Borg, 1996), even in consonant clusters with voiced (e.g., *qdart* /ʔdɑrt/ English, 'I dared' and *bqajt* /bʔajt/ English, 'I remained') and voiceless stops (e.g., *qtates* /ʔtates/, English 'cats' and *tqaqpiq* /tʔɑʔpɪʔ/, English 'honking of a car horn'). Nevertheless, it also can occur as an epenthetic segment to mark otherwise vowel-initial words (e.g., *attur* – /at:ur/ →[ʔat:ur]) (Mitterer, Kim, & Cho, 2019), as is well-known for English, Dutch, and German (e.g., *the eagle*, /ðə#iːgˠl/ → [ðəʔiːgˠl]). Mitterer et al. (2019) found that such epenthetic glottal stops are employed in roughly 50% of the eligible cases, and that epenthetic glottal stops do, on average, not differ from underlying glottal stops in terms of their phonetic properties. Moreover, epenthetic glottal stops are more likely when speakers inserted a small prosodic boundary before the (otherwise) vowel-initial word. This was apparent, because, in phrases such as *il-kliem attur u dar* (English, 'the words actor and house'), an epenthetic glottal stop on the word *attur* /at:ur/ was more likely when the preceding word had some final lengthening. This suggests a prosodic function of the epenthetic glottal stop. As such, we would also expect that an epenthetic glottal stop is more likely if the (otherwise) vowel-initial words is under contrastive focus, since epenthetic glottal stops have also been associated with prominent syllables in other languages (Garellek, 2014; Jongenburger & van Heuven, 1991). Since we assume that listeners have stored a glottal stop-initial version of the vowel-initial words (Mitterer et al., 2019), this phonological process does not necessarily impede word recognition. Moreover, Garellek (2013) argued that glottalization enhances high-frequency energy in the following vowel, which may enhance the spectral distinctiveness between vowels. As such, it would not impair the performance when a vowel-initial word is contrasted with a lexical competitor (i.e., when the word *actor* [*attur*] contrasts with the word *theatre* [*teatru*]) and, given the cross-language tendency to associate glottal stops with prominence (Davidson, 2021), would lend prominence to this word. However, in case of a particular phonological contrast, adding the glottal stop would be counterproductive. In case of the pair *attur* /at:ur/ – *qattus* /ʔat:us/, adding an epenthetic glottal stop to *attur* [ʔat:ur] in fact makes the words more similar rather than distinct. That is, in this case, the speaker would be well advised to not produce an epenthetic glottal stop; Mitterer et al. (2019) found that words produced with an epenthetic glottal stop are in fact strong competitors for words with underlying glottal stops. That is, in a visual-world eye-tracking task, participants looked at words such as *qattus* if the target was *attur* that happened to be produced with an epenthetic glottal stop. This suggests that the epenthetic glottal stop makes it more difficult for listeners to distinguish the pair *attur* /at:ur/ – *qattus* /ʔat:us/, so that speakers would be well advised to produce the vowel-initial words as such (i.e., without an epenthetic glottal stop) to maximize the contrast. This is the question we investigated in this paper by asking participants

to produce vowel-initial words such as *attur* in a sentence as given information—leading to a reduced pronunciation with a shorter duration and little f0 movement—or under either lexical or phonological competition—leading to an acoustically prominent pronunciation with a longer duration and more f0 movement. The phonological competitor was a word that started similarly as target word but had an underlying glottal stop (such as *qattur* for *attus*). The prediction of a listener-oriented account would have to be that speakers do not use the epenthetic glottal stop in this phonological-contrast condition to maximize the contrast between the target word and the word it contrasts with.

Before delving into the method, it is worthwhile to consider how information structure may influence phrasing in Maltese. Maltese expresses focus using a pitch accent on the stressed syllable of the focussed constituent (Grice, Vella, & Bruggeman, 2019; see in particular Figures 4.12 and 4.13 in Vella, 1994). Lexical stress is assigned on phonological grounds in Maltese, mostly falling on the penultimate syllable, unless the final syllable contains a geminate consonant or vowel (Grice et al., 2019). Importantly, Vella and colleagues (Grice et al., 2019; Vella, 1994, 2003) do not mention that Maltese may use a pause before focussed material to highlight information structure. This is crucial for the current project, since we need speakers to speak fluently, otherwise, it is very difficult to determine whether a vowel-initial word is produced with an initial glottal stop or not, since silence is the clearest sign of a full glottal stop. To investigate whether speakers nevertheless introduce a prosodic boundary before the contrasting material—despite the instruction to produce the sentences fluently, we measured the duration of the word preceding the critical word and tested whether there is a slow-down—indicative of a prosodic break—before the critical word. If speakers introduce a prosodic break before the critical word, we should find that the preceding word is longer, indicative of phrase-final lengthening when there is a contrast to be made.

2. Method

2.1. Participants

A total of 24 participants, ages ranging from 18 to 33 (M = 21.71, *SD* = 3.93) participated in the study, and were financially compensated for their time. From this total, one participant did not manage to finish the session, as he was not able to produce the sentences fluently. Therefore, the experimental session was aborted. Of the remaining 23 participants, 16 were male with ages ranging from 19 to 29 (M = 21, *SD* = 2.58) and seven were female, ages ranging from 18 to 33 (M = 23.29, *SD* = 6.18). Given that Malta is officially bilingual (Maltese and [Maltese] English), we asked participants about their language use during their lifetime. That is, they had to indicate for various parts of their lives, which language they used to which degree. **Table 1** shows the data, indicating that Maltese was the primary language for most participants.

	Language					
	Maltese			English		
	Mean	**Median**	*SD*	**Mean**	**Median**	*SD*
Infancy 0–6	79.1	90	17.6	20.2	10	17.0
Childhood 7–12	71.5	70	18.3	26.7	20	17.5
Adolescence 13–18	65,0	65	19.0	29.8	30	16.8
Adulthood (18+)	62.8	60	18.3	32.4	30	17.4

Table 1: Self-reported percentages of use of language at different ages. Values provide the median, mean, and standard deviation.
Note: The missing percentages to 100% are mostly taken up by Italian as another language often used on Malta.

2.2. Apparatus and stimuli

This experiment made use of a sound-attenuated booth at the Cognitive-Science lab facility of the University of Malta. The experiment was conducted on a standard PC running Speechrecorder (Draxler & Jänsch, 2004). All the visual materials were displayed from a standard monitor running at a refresh rate of 60 Hz. Participants were recorded using a Focusrite CM25 large diaphragm condenser microphone connected to a Focusrite 2i2 USB audio interface for D/A conversion.

The stimuli materials used in this experiment were based on the items that were used in the Mitterer et al. (2019) experiment. From their list of 34 pairs, we focussed on 6 pairs of words in which there is sufficiently long (pseudo) overlap at the beginning of the word (such as *qattus-attur*; see the appendix for the full list) with additionally the same number of syllables and the same stress pattern. From these pairs, stimuli using the question-answer format used by Cho et al. (2017) were generated which put the target word either under phonological contrast, semantic contrast, or no contrast for an unaccented baseline. Visual prompts were created for these three conditions in two sentence frames (see **Table 2**), in which the question and answer was provided with the contrasting word highlighted in red ink. To focus on the contrast between vowel-initial words and glottal stop initial words, the glottal stop initial words were also used as targets. This leads to 12 targets in six possible question-answer pairs, leading to a total of 72 different prompts.

2.3. Procedure

The participants were instructed to read aloud the answer, with emphasis on the highlighted word but without pauses. It was ensured that all participants were aware of this specific protocol, and whenever the researcher noticed any different response, the participants were reminded of the importance to emphasize the highlighted word while maintaining fluency in the utterance.

Condition	Questions and Answers	Contrast
Frame 1		
Question	Q: *Nina fehmet attur f'dan il-każ?* Q: 'Did Nina understand actor in this case?'	
Answer/Given	A: *Le, Anna fehmet attur f'dan il-każ* A: 'No, Anna understood actor in this case'	Nina vs. Anna
Answer/Phonological competitor	*Anna fehmet qattus f'dan il-każ?* 'Did Anna understand cat in this case?'	attur vs. qattus
Answer/Semantic competitor	*Anna fehmet teatru f'dan il-każ?* 'Did Anna understand theatre in this case?'	attur vs. teatru
Frame 2		
Question	Q: *Ir-riposti t-tajbin ġew attur u post?* Q: 'The correct answer was actor and place?'	
Answer/Given	A: *Le, ir-risposti t-tajbin/l- ħżiena ġew attur u post* A: 'No, the wrong answer was actor and place'	correct vs. wrong
Answer/Phonological competitor	*Ir-riposti t-tajbin ġew qattus u post?* 'The correct answer was cat and place?'	attur vs. qattus
Answer/Semantic competitor	*Ir-riposti t-tajbin ġew teatru u post?* 'The correct answer was theatre and place?'	attur vs. teatru

Table 2: Questions to elicit unaccented and contrastive versions of the word 'attur.'
Note: The answer is the same for all three cases of both examples.

Each participant was exposed to a total of 288 critical trials. Each of the 12 possible targets was used in six possible question-answer trials, leading to a total 72 prompts per block. This was then repeated four times in four different blocks, for a total of 288 trials. Within blocks, the order of presentation was randomized. Prior to starting the experiment, the participants were allowed to practice using printed examples outside the sound-attenuated booth to allow the researcher to provide verbal instructions. After the participants were comfortable with the task, they proceeded to the actual experiment, which started with twelve practice trials. Given the lengthy nature of the experiment, the participants were made aware that they can pause at any time desired to rest.

As described above, the nature of the experiment required the participants to skim through the question and then read aloud the answer, with an accent on the highlighted word. For the *semantic-* and *phonological-competitor* condition, the highlighted word would be the target word (e.g., didwi, għalqa, etc.), but for the *given* condition, the highlighted word would be a different one than the target (see **Table 2**).

2.4. Data analysis

The recordings were pre-processed as follows. Starting off, the resulting sound files from the 23 participants were forced-aligned using MAUS web interface (Kisler, Reichel, & Schiel, 2017). The forced-alignment included a rule that any word-initial vowel may have an epenthetic glottal stop. After the algorithm was run for all entries, each utterance and its alignment were hand-checked and marked as either usable, disfluent in the target region (i.e., the target word and its preceding word), or with a clearly invalid forced alignment. Failures of forced alignment, for instance, occurred in recordings in which the participants produced a fluent answer, but there was an initial breathing noise that was then aligned to be the *'Le'* of the utterance, leading to a forced alignment that was clearly off the mark for the rest of the sentence. These recordings were trimmed so that there were no long initial or final silences or extraneous noises. After trimming, the resulting sound files were again subjected to forced alignment. This led to a good alignment on all cases in which the participant provided a fluent answer. In total, only 109 trials (1.64%) were rejected for non-fluent productions in the target regions.

As observed in Mitterer et al. (2019), the forced-alignment algorithm is not able to notice glottalization. The phone model for the glottal stop apparently requires a full closure to be triggered. Moreover, segments have a minimum duration of 30 ms in the forced-alignment algorithm, so that shorter silences would go unnoticed. That is, the forced-alignment algorithm has a fairly high threshold for detecting a glottal stop, so that all alignments without a glottal stop require manual inspection for signs of glottalization before one can conclude that there is no glottalization. This was achieved using the same criteria as used in Mitterer et al. (2019). As reported by Redi and Shattuck-Hufnagel (2001), glottalizations may come in different shapes, such as drop or discontinuity in the f0 contour, or a drop in the amplitude contour. Mitterer et al. (2019) provide examples of such cases in their appendix. If such a signal, that is, a discontinuity in amplitude or f0 contour, was found, an epenthetic glottal stop was manually added to the forced alignment. **Figure 1** provides examples of a full glottal stop as inserted in the forced-alignment procedure (i.e., by the rule V → ʔV/#V), a glottalization as inserted during hand correction, and a fluent production with no glottal gesture.

For each item that could be analyzed, we extracted whether there is a glottal gesture at the beginning of the word, as well as three prosodic indicators we expect to vary with the presence of an accent: mean and standard deviation of the f0 contour, and target-word duration. Based on previous research (see Watson, 2010, for a review), we can expect that these measures are larger when a word is contrasted with a competitor, so that these measures help us to determine whether the participants indeed produce an accent on these words. Note that we use the standard deviation of the f0 contour as a measure of how much f0 movement there is. Often f0 range is used, but the standard deviation is more robust regarding outliers.
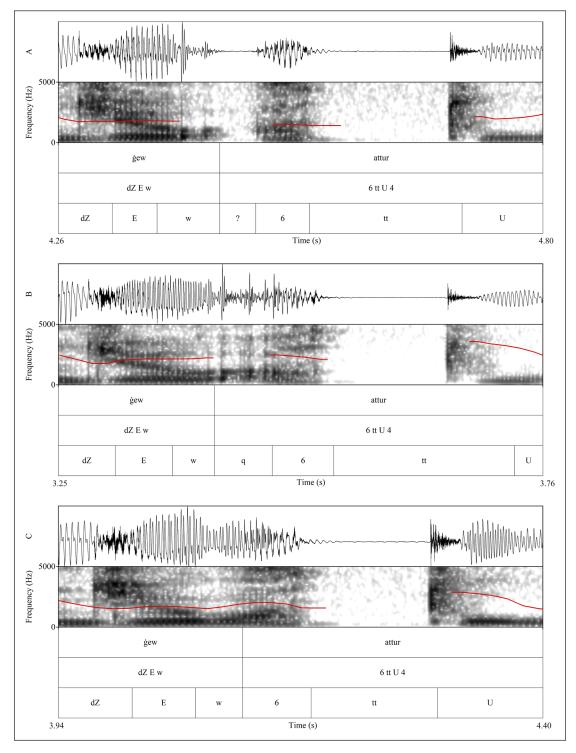
Figure 1: Three Examples of the Phrase 'ġew attur' with a Full Glottal Stop **(A)**, Glottalization **(B)**, or no glottal marking **(C)**. The red line shows the pitch tracking.

These data were then analyzed with linear mixed effect models. To estimate significance, the Satterthwaite's degrees of freedom method was used as implemented in the package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015). For the (categorical) main independent variable, a generalized version of the linear mixed effect model with a logit-link was used. The models were run with the independent variables Repetition and Contrast condition. This three-level condition contrast was contrast-coded into two contrasts. Given versus Contrasting (contrast weights: given: –2/3, semantic and phonological competitor: 1/3), and type of contrast (contrast weights: given 0, semantic competitor –0.5, phonological competitor 0.5). The first of these contrasts hence tests whether there is a difference between the given condition compared to both contrasting conditions. The second contrast tests whether there is a difference between the two types of contrast. All four models started off with a full random-effect model and random effects were simplified (i.e., random slopes removed) until convergence was reached (see the R markdown file on OSF for details).

3. Results

Prior to starting the analysis, the data from one participant was disregarded due to disfluencies in the majority of utterances, including either the target or preceding words. For the remaining data, we first evaluated whether the experiment was successful in inducing an accent if there is a contrast and then the main question of how the different types of contrast influenced the use of the epenthetic glottal stop. The data and the analysis code are available at: https://osf.io/8ejh3/.

3.1. Indicators of contrast

Figure 2 shows that under contrast, words are produced longer, with a higher mean f0 and a higher standard deviation of the f0 contour for the competitor conditions compared to the given condition. This shows that the experiment successfully made speakers produce the target word more prominently. The results also show little effect of what type of contrast is putting the target word under focus (i.e., semantic versus phonological).

These impressions are borne out by statistical testing. All models (see data analysis section for details) investigated whether there was an effect of contrast and repetition on the three dependent variables mean f0, standard deviation f0, and word duration. All three models (see **Tables 3–5**) show a clear difference between the two contrasting conditions and the given condition. That is, these results indicate that participants consistently accented the target words when it was under contrast. No differences were found between the semantic and the phonological competitor condition.
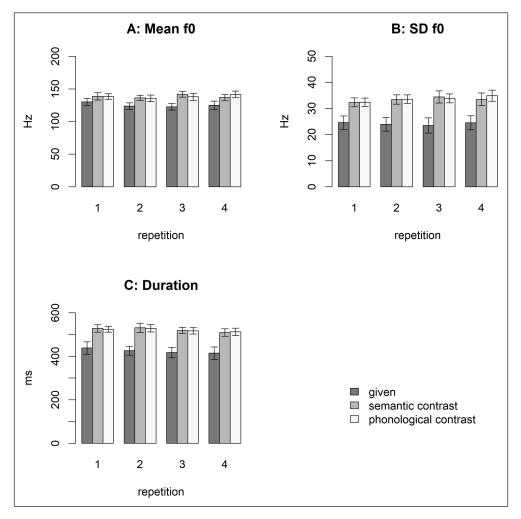
Figure 2: The Effect of Condition (Given, Semantic Contrast, and Phonological Competitor) and Repetition on Mean f0 **(Panel A)**, f0 Standard Deviation **(Panel B)**, and Word Duration **(Panel C)** of the Target Word. Error bars show the within-participant confidence intervals (following Morey, 2008).

Predictor	B (SE)	t (df)	*p*
Intercept	134.002 (7.08)	18.936 (23)	<0.001
contrasted vs. given	13.012 (3.25)	4.007 (16)	0.001
semantic vs. phonological competitor	0.04 (1.76)	0.023 (5)	0.983
repetition	–0.135 (0.74)	–0.182 (22)	0.857
contrasted vs. given: repetition	2.212 (1.21)	1.834 (3162)	0.067
semantic vs. phonological competitor: repetition	–1.22 (1.39)	–0.88 (3161)	0.379

Table 3: Results of the Linear Mixed-Effect Model for the Mean f0 Contour of the Target Words.

Predictor	B (SE)	t (df)	p
Intercept	30.422 (2.14)	14.186 (26)	<0.001
contrasted vs. given	9.413 (1.54)	6.116 (22)	<0.001
semantic vs. phonological competitor	–0.258 (0.55)	–0.47 (3144)	0.638
repetition	0.39 (0.3)	1.305 (22)	0.205
contrasted vs. given: repetition	0.681 (0.5)	1.369 (22)	0.185
semantic vs. phonological competitor: repetition	–0.334 (0.49)	–0.679 (3144)	0.497

Table 4: Results of the Linear Mixed-Effect Model for the Standard Deviation of the f0 Contour over the Target Words.

Predictor	B (SE)	t (df)	p
Intercept	488.578 (21.25)	22.987 (20)	<0.001
contrasted vs. given	97.266 (15.95)	6.099 (23)	<0.001
semantic vs. phonological competitor	0.804 (3.18)	0.253 (3139)	0.801
repetition	–6.411 (2.93)	–2.189 (20)	0.041
contrasted vs. given: repetition	2.371 (3.54)	0.67 (22)	0.51
semantic vs. phonological competitor: repetition	–2.288 (2.84)	–0.804 (3139)	0.421

Table 5: Results of the Linear Mixed-Effect Model for the Duration of the Target Words.

3.2. Glottalization of vowel-initial words

Figure 3 shows the proportion of glottalizations observed for each condition over the course of the experiment. Overall, there are about 50% of epenthetic glottal stops in the *given* condition. This is in line with the results reported in Mitterer et al. (2019), which made use of a similar elicitation task in which the vowel-initial words were given.

As expected, contrasting target words more often contain an epenthetic glottal stop than given target words. In the phonological-contrast condition, the proportion of epenthetic glottal stops is slightly higher than in the semantic-contrast condition.

The analysis (see **Table 6**) showed that there is a clear difference between the given condition and the two conditions with a competitor, but also a smaller, but significant difference between the semantic- and phonological-competitor condition. The latter may be surprising given the overlap in confidence intervals in **Figure 3**, but it has to be noted that these confidence intervals are for each repetition separately, while the analysis reveals, when taking into account the whole data set, there is a difference between the two contrast conditions. Remarkably, this difference is in the opposite direction of what a communicative account predicts, with more epenthetic glottal stops in the phonological-competitor condition than in the semantic-competitor condition. While
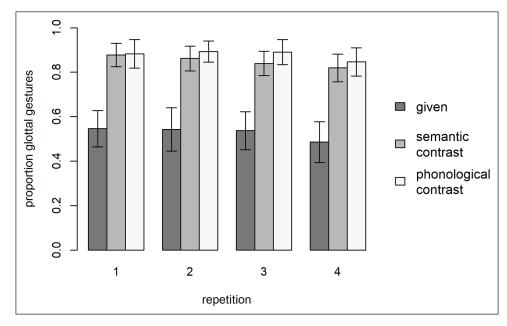
Figure 3: The Effect of Condition (Given, Semantic-Contrast, and Phonological-Contrast) and Repetition on the likelihood of Glottalization of Vowel-Initial Words. Error bars show the within-participant confidence intervals (following Morey, 2008).

	B (*SE*)	***z***	***p***
Intercept	1.616 (0.319)	5.068	<0.001
contrasted vs. given	2.242 (0.106)	21.123	<0.001
semantic vs. phonological competitor	−0.304 (0.142)	−2.145	0.032
repetition	−0.132 (0.078)	−1.679	0.093
contrasted vs. given: repetition	−0.025 (0.092)	−0.267	0.789
semantic vs. phonological competitor: repetition	−0.079 (0.126)	−0.623	0.534

Table 6: Results of the Linear Mixed-Effect Model for the Presence of a Glottal Gesture.

the difference is small in absolute terms (84.9% with a semantic competitor and 87.7% with a phonological competitor), this shows that the experiment was able to pick up small differences, and it makes it unlikely that the absence of support for a listener-oriented account is due to a lack of statistical power.

3.3. Duration of the word preceding the target

We performed additional analyses on the duration of the previous word to see whether participants are trying to introduce a prosodic boundary before a potentially prominent word. This would be another rationale for introducing an epenthetic glottal stop (Mitterer et al., 2019). We predicted the duration of the word preceding the critical target word using the predictor

Repetition, the same contrast-coded predictors for information structure (*given* versus *contrasted* and *phonological* versus *semantic competitor*), and the identity of the preceding word (*il-kliem* versus *ġew*, given that the word with fewer segments is highly likely to be shorter). We performed separate analysis for the vowel-initial items and the glottal stop-initial items, because, for the vowel-initial items, we also took into account whether there is an epenthetic glottal stop, and this predictor cannot be used for the glottal stop-initial items. The results (see **Table 7**) show no lengthening due to the information structure (i.e., the predictors *contrasted* versus *given* and *semantic* versus *phonological competitor*). However, there is an effect of glottal-stop insertion on the duration of the preceding word.

This might be caused by the fact that glottal-stop insertion might be related to overall speaking rate or after all reflect that the preceding word is longer if it is followed by an accented word. These two accounts can be contrasted by analyzing the duration of the preceding word separately for the given and the contrasted conditions. If this effect is due to the tendency to insert a boundary before a contrasted word, the effect should be absent for the given words but clear for the contrasted words. If this effect is related to speech rate, it should be present in both cases. The results are in line with the latter prediction, as the separate analyses show an effect of glottal-stop insertion on the duration of the preceding word for both contrasted (b = 18.3, SE = 7.05), t(19.84) = 2.608, p = 0.0169) and given target words (b = 33.1, SE = 6.093, t(20.47) = 5.439, p < 0.001). This shows that utterances with an epenthetic glottal stop may generally be slower than those without, independent of whether the critical word bears an accent or not.

	B (*SE*)	t(df)	*p*
Glottal-stop initial words			
Intercept	311.124 (8.94)	34.782 (27)	<0.001
Version (*il-kliem* vs. *ġew*)	172.623 (1.67)	103.295 (3214)	<0.001
Repetition	–4.76 (0.75)	–6.373 (3214)	<0.001
contrasted vs. given	0.927 (2.04)	0.454 (3214)	0.65
semantic vs. phonological competitor	4.735 (4.51)	1.049 (22)	0.305
Vowel-initial words			
Intercept	301 (10)	30.718 (25)	<0.001
Version (*il-kliem* vs. *ġew*)	177 (1.6)	1.6024 (177)	<0.001
Repetition	–5 (0.7)	–6.438 (3181)	<0.001
contrasted vs. given	–1 (1.9)	–0.495 (3161)	0.62
semantic vs. phonological competitor	5 (3.3)	1.374 (20)	0.185
+ epenthetic glottal stop	26 (6.5)	4.085 (19)	<0.001

Table 7: Results of the Linear Mixed-Effect Models for the duration of the preceding word.

4. Discussion

In this experiment, we investigated how epenthetic glottal stops in Maltese behave under three conditions, when the vowel-initial word potentially triggering epenthesis was given and when it should be differentiated from a semantic or phonological competitor (phonological, e.g., *qattus* versus *attur,* English 'cat' versus 'actor' versus semantic, *teatru* versus *attur,* English 'theatre' versus 'actor'). A previous study by Cho et al. (2017) had shown that listeners may use strengthening strategies that may in fact run counter to the prediction of a listener-oriented account, so that the strengthening in fact makes the word more rather than less similar to its phonological competitor. This finding contrasted with other findings that speakers are able to selectively enhance critical cues to enhance the difference between a target word and a given phonological competitor (Buz et al., 2016; Schertz, 2013; Seyfarth et al., 2016) in a listener-oriented way.

We asked speakers to make a contrast between words such as *attur* and *qattus.* Usually, such contrasts lead speakers to produce the target word more prominently, which in turn might lead to the introduction of an epenthetic glottal stop for the vowel-initial word *attur* (Davidson, 2021; Garellek, 2013). Under the assumption of audience design, such a phonological contrast should, however, lead to a reduction of epenthetic glottal stops, as this maximizes the phonological distance between the target and its phonological competitor. In contrast with this prediction, we found that vowel-initial words that contrasted phonologically with their antecedent had even more glottal markings than vowel-initial words that contrasted semantically with their antecedent.

We investigated whether the epenthetic glottalization might be a consequence of lengthening prior to an accented constituent. While we found that the preceding words were not longer in case of a semantic or phonological contrast, we found that utterances with an epenthetic glottal stop seem to be generally longer. This may be surprising, since Seyfarth and Garellek (2020) found that /t/ in mid-west American English is more likely to be glottalized at faster rates. While it may seem contradictory that one finds more glottalization at lower speech rates in Maltese but less glottalization at lower speech rates in English, the findings in fact dovetail well with each other. In English, /t/ glottalization is a reduction process while the insertion of an additional phonetic element (such as glottal-stop insertion) is an enhancement process, so one may even expect opposite effects.

One might wonder whether the absence of a listener-oriented effect for glottal-stop insertion is due to the fact that the word pairs were only (pseudo-)cohort competitors and not full minimal pairs. This may lead speakers to not put in a full effort to distinguish the words, since they will be distinguished clearly at word offset. This post-hoc assumption is, however, difficult to maintain since we used the epenthetic glottal stop in Maltese to replicate a similar version of such a counterproductive pattern as observed in Cho et al. (2017) in a different language and with a different phonetic marker. Cho et al. (2017) had used minimal pairs (*e.g., bomb* versus *Bob*) and

obtained similar results. This makes it difficult to maintain that our results are due to the use of cohort competitors rather than minimal pairs.

Cho et al. (2017) found that both types of contrast led to similar effects in terms of articulatory strengthening. Our data, in contrast, show a difference between the semantic- and phonological-contrast condition. This indicates that speakers may indeed try to establish these contrasts differently. However, the means to do that was in fact counterproductive by producing more epenthetic glottal stops in the phonological-contrast condition, because this reduced the difference between the target word and its phonological competitor. We can only speculate how to explain this difference. Watson (2010) argued that prominence is not a unitary phenomenon but may vary continuously. Based on this assumption, it is possible to argue that a phonological contrast might be stronger than a lexical contrast, and since glottal stop insertion is related to prominence (Garellek, 2013), this may lead to more epenthesis. It remains unclear though, why similar results were then not also found by Cho et al. (2017) in their study of nasal coarticulation, but maybe the effect might be stronger (and hence easier to find) for glottal-stop insertion than for reducing nasal coarticulation.

Given that our findings fail to support the assumption that speakers take the listeners' perspective fully into account on a phonetic level, how to account for the results reviewed in the introduction that did support this assumption (Buz et al., 2016; Schertz, 2013; Seyfarth et al., 2016)? An obvious candidate is that participants in those studies believed they were interacting with an interlocutor or a speech recognizer. Maybe the absence of an interactive task made participants in this experiment (and in Cho et al., 2017) use suboptimal strategies? There are three reasons to be skeptical about this explanation. First of all, studies that compared production with and without the presence of an interlocutor often provided similar patterns of results (Baese-Berk & Goldrick, 2009; Ferreira & Dell, 2000; Fox, Reilly, & Blumstein, 2015). Secondly, participants clearly made the target words more prominent under contrastive focus in our experiments, that is, they engaged in the task of generating contrasts. Third, such an explanation leaves unexplained one detail in the results of Schertz (2013). In that study, no consistent enhancement of vowel-quality differences was found.

This does not mean, however, that speakers in general cannot enhance vowel-quality differences. However, such changes when speaking clearly are relatively blunt, by enlarging the vowel space in general (Ferguson & Kewley, 2007; Hazan & Baker, 2011). At this stage it is important to note that it was never the question whether speakers can modify vowel properties in a listener-oriented way. In fact, the data of Cho et al. (2017)—which partly inspired the current study—had already shown that listeners even modulate fine phonetic details such as nasal coarticulation on vowels. The issue is therefore not whether vowel properties can be modulated; the question is whether vowel properties can be modulated in a way that fully takes the listener

into account in all cases. It is important to mention that the findings of Cho et al. (2017) indicate that speakers, when trying to convey a phonological contrast, try to make each segment clearer by reducing coarticulatory influences. More often than not, this is a useful strategy, for instance, when trying to convey that the intended word was *Ben* rather than *ban* or *bin*. However, in the odd case (here: when the contrasting word is *bed*), this strategy is not useful. Similarly, in the current case, we are not arguing that the insertion of the glottal stop *per se*, is problematic, because it might lead to a phonological variant form. Our earlier work (Mitterer et al., 2019) has shown that the form of a vowel-initial word with an epenthetic glottal stop is quite well recognized. There is hence by default no cost associated with the epenthetic glottal stop, possibly, because as Mitterer et al. (2019) argued, listeners have stored the glottal-stop initial form of vowel initial words in their mental lexicon. The epenthetic glottal stop may in general even make the word more easily recognizable by the additional marking of the word boundary. This may sound odd, because Maltese listeners—in contrast to English listeners—cannot take a glottal-stop to be a boundary marker since it might simply be another segment. However, adding the glottal stop creates additional time for the lexical competition from the previous word to settle, which is a major cue for a word boundary (Mattys, White, & Melhorn, 2005). So, as in Cho et al. (2017), glottal-stop epenthesis may be a strategy that, in principle may help the listener, however, not in the case when the contrasting word is one that also starts with a glottal stop. That is, we argue that speakers have strategies at their disposal that in general enhance the speech signal for the listener, but those mechanisms are fairly blunt—though usually effective.

This account seems to be in contrast with the data of Turnbull et al. (2018), who found that nasal place assimilation is constrained in a way that seems listener-oriented in a rather specific way, with predictable words being assimilated and unpredictable following words triggering assimilation. Turnbull et al. (2018) discuss a potential prosodic account which focusses on the temporal overlap of the gestures. This account is, as they discuss, indeed unlikely. However, an alternative account for how this behavior could fall out of the speech-production process would be using prosodic weight, which may be influenced by the difficulty to retrieve the word during the formulation process. That is, if a word is difficult to retrieve, it receives more prosodic weight, which makes it less susceptible for contextual influences, such as phonological assimilation. It is important to note that such a process may in fact lead to a listener-oriented outcome, that is, speakers learn that words that are difficult to retrieve may require a clearer pronunciation. Such a mechanism would, however, take the burden away from the speaker to estimate, for each parameter, how it would have to be produced. That is, it would not require speakers to estimate the exact phonetic effects, similar to our own findings, in which listeners appear to try to be even clearer for a phonological contrast compared to a lexical one but fail to implement the phonetically best strategy given the contrast in question. Such a mechanism would also not be well described as 'availability based,' which is usually contrasted with a listener-oriented

strategy. Low availability should only lead to slower pronunciation leading up to the word in question, or the insertion of other material (such as optional words or fillers), but not of a clearer pronunciation of the difficult-to-retrieve word itself and increased resistance to coarticulation in case of slow retrieval. That is, we argue that the basis of this mechanism may in fact be listener-oriented, even though it does not lead to an ideal outcome in all cases.

It is beyond the scope of the current paper to fully unite all findings when speakers are able to specifically enhance a given contrast and when not. However, one might speculate whether 'phonological awareness,' which is usually linked to orthographic coding (Morais, 2021) may play a role. In another study on epenthetic glottal stops in Maltese (Mitterer, Kim, & Cho, 2020), we observed that epenthetic glottal stops influenced parsing decision, but participants were not able to verbalize that they were influenced by an epenthetic glottal stop and never reported hearing such segments in a post-experiment interview. Obviously, measuring awareness is difficult, and a post-experiment interview may be too late to find evidence of a fleeting awareness. Nevertheless, it is possible that phonological awareness may contribute to the ability to specifically enhance a given phonological contrast, with voicing contrasts being easy and epenthetic glottal stops being difficult. This would be an avenue for further research.

In sum, our data indicate that speakers may not always be fully attuned to the listeners' needs when trying to strengthen a phonological contrast. In contrast to voicing differences in stops (Schertz, 2013), coarticulatory nasalization (Cho et al., 2017) and epenthetic glottal stops may be more difficult to manipulate in a situation-specific way. This indicates a clear limit to the extent that speakers can take listeners into account.

Additional file

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

- **Appendix.** The six target words and their lexical and phonological contrasts. DOI: https://doi.org/10.5334/labphon.6441.s1

Competing interests

The authors have no competing interests to declare.

References

Aylett, M., & Turk, A. (2016). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*. DOI: https://doi.org/10.1177/00238309040470010201

Azzopardi-Alexander, M., & Borg, A. (1996). *Maltese* (1 edition). Routledge.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554. DOI: https://doi.org/10.1080/01690960802299378

Burdin, R. S., Turnbull, R., & Clopper, C. G. (2014). Interactions among lexical and discourse characteristics in vowel production. *Proceedings of Meetings on Acoustics*, *22*(1), 060005. DOI: https://doi.org/10.1121/2.0000084

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68–86. DOI: https://doi.org/10.1016/j.jml.2015.12.009

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

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678. DOI: https://doi.org/10.1121/1.1810292

Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *WIREs Cognitive Science*, *12*(3), e1547. DOI: https://doi.org/10.1002/wcs.1547

de Boer, B. (2000). Self-organisation in vowel systems. *Journal of Phonetics*, *28*, 441–465. DOI: https://doi.org/10.1006/jpho.2000.0125

Draxler, C., & Jänsch, K. (2004). Speechrecorder—A universal platform independent multi-channel audio recording software. *Proceedings of LREC*, 559–562. DOI: https://doi.org/10.1016/j.lingua.2012.12.006

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41. DOI: https://doi.org/10.1044/1092-4388(2007/087)

Ferguson, S. H., & Kewley, -Port Diane. (2007). Talker Differences in Clear and Conversational Speech: Acoustic Characteristics of Vowels. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1241–1255. DOI: https://doi.org/10.1044/1092-4388(2007/087)

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

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language, 83,* 97–117. DOI: https://doi.org/10.1016/j.jml.2015.04.002

Garellek, M. (2013). *Production and perception of glottal stops* (Doctoral dissertation, UCLA). https://escholarship.org/uc/item/7zk830cm

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

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

Grice, M., Vella, A., & Bruggeman, A. (2019). Stress, pitch accent, and beyond: Intonation in Maltese questions. *Journal of Phonetics, 76,* 100913. DOI: https://doi.org/10.1016/j.wocn.2019.100913

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America, 130*(4), 2139–2152. DOI: https://doi.org/10.1121/1.3623753

Hura, S. L., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35,* 59–72. DOI: https://doi.org/10.1177/002383099203500206

Jaeger, T. F., & Buz, E. (2018). Signal reduction and linguistic encoding. In *The handbook of psycholinguistics* (pp. 38–81). Wiley Blackwell. DOI: https://doi.org/10.1002/9781118829516.ch3

Jongenburger, W., & van Heuven, V. J. (1991). The distribution of (word initial) glottal stop in Dutch. *Linguistics in the Netherlands, 8*(1), 101–110. DOI: https://doi.org/10.1075/avt.8.13jon

Katseff, S., Houde, J., & Johnson, K. (2012). Partial Compensation for Altered Auditory Feedback: A Tradeoff with Somatosensory Feedback? *Language and Speech, 55*(2), 295–308. DOI: https://doi.org/10.1177/0023830911417802

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language, 45,* 326–347. DOI: https://doi.org/10.1016/j.csl.2017.01.005

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*(2), 194–231. DOI: https://doi.org/10.1016/j.cogpsych.2004.08.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R Package Version, 2.*

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9*(3), 550–557. DOI: https://doi.org/10.3758/BF03196312

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology. General, 134*(4), 477–500. DOI: https://doi.org/10.1037/0096-3445.134.4.477

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. DOI: https://doi.org/10.1016/j.jml.2019.104034

Mitterer, H., Kim, S., & Cho, T. (2020). The Role of Segmental Information in Syntactic Processing Through the Syntax–Prosody Interface. *Language and Speech*, 0023830920974401. DOI: https://doi.org/10.1177/0023830920974401

Morais, J. (2021). The phoneme: A conceptual heritage from alphabetic literacy. *Cognition, 213*, 104740. DOI: https://doi.org/10.1016/j.cognition.2021.104740

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. *Tutorials in Quantitative Methods for Psychology, 4*, 61–64. DOI: https://doi.org/10.20982/tqmp.04.2.p061

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech (2nd release)*. Department of Psychology, Ohio State University (Distributor).

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407–429. DOI: https://doi.org/10.1006/jpho.2001.0145

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3), 249–263. DOI: https://doi.org/10.1016/j.wocn.2013.03.007

Schwartz, J.-L., Boë, L. J., Valleé, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of Phonetics*, *25*, 233–253. DOI: https://doi.org/10.1006/jpho.1997.0044

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *The Journal of the Acoustical Society of America, 139*(2), EL31–EL37. DOI: https://doi.org/10.1121/1.4942544

Seyfarth, S., & Garellek, M. (2020). Physical and phonological causes of coda /t/ glottalization in the mainstream American English of central Ohio. *Laboratory Phonology*, *11*(1), Article 1. DOI: https://doi.org/10.5334/labphon.213

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, *3*(1), 236–264. DOI: https://doi.org/10.1111/j.1749-818X.2008.00112.x

Steriade, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219–250). Academic Press.

Turnbull, R. (2018). Patterns of probabilistic segment deletion/reduction in English and Japanese. *Linguistics Vanguard*, *4*(s2). DOI: https://doi.org/10.1515/lingvan-2017-0033

Turnbull, R., Seyfarth, S., Hume, E., & Jaeger, T. F. (2018). Nasal place assimilation trades off inferrability of both target and trigger words. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 15. DOI: https://doi.org/10.5334/labphon.119

Vella, A. (1994). *Prosodic structure and intonation in Maltese and its influence on Maltese English.*

Vella, A. (2003). Phrase accents in Maltese: Distribution and realization. *Proccedings of ICPhS*. 15th International Congress of Phonetic Sciences, Barcelona., Barcelona.

Watson, D. G. (2010). The Many Roads to Prominence: Understanding Emphasis in Conversation. In *Psychology of Learning and Motivation, 52*, 163–183. Academic Press. DOI: https://doi.org/10.1016/S0079-7421(10)52004-8