

## Word final schwa is driven by intonation—The case of Bari Italian

Martine Grice, Michelina Savino, and Timo B. Roettger

Citation: *The Journal of the Acoustical Society of America* **143**, 2474 (2018); doi: 10.1121/1.5030923

View online: <https://doi.org/10.1121/1.5030923>

View Table of Contents: <http://asa.scitation.org/toc/jas/143/4>

Published by the [Acoustical Society of America](#)

---

### Articles you may be interested in

[Recoverability-driven coarticulation: Acoustic evidence from Japanese high vowel devoicing](#)

*The Journal of the Acoustical Society of America* **143**, 1159 (2018); 10.1121/1.5024893

[Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker](#)

*The Journal of the Acoustical Society of America* **143**, 2013 (2018); 10.1121/1.5027410

[Speaking rhythmically improves speech recognition under “cocktail-party” conditions](#)

*The Journal of the Acoustical Society of America* **143**, EL255 (2018); 10.1121/1.5030518

[Discovering acoustic structure of novel sounds](#)

*The Journal of the Acoustical Society of America* **143**, 2460 (2018); 10.1121/1.5031018

---

# Word final schwa is driven by intonation—The case of Bari Italian

Martine Grice<sup>a)</sup>

*University of Cologne, IfL-Phonetik, Herbert-Lewin-Strausse 6, Köln, 50931, Germany*

Michelina Savino

*University of Bari “Aldo Moro,” Department of Education, Psychology, Communication, Piazza Umberto I, 1, Bari, 70121, Italy*

Timo B. Roettger<sup>b)</sup>

*University of Cologne, IfL-Phonetik, Herbert-Lewin-Strausse 6, Köln, 50931, Germany*

(Received 16 October 2017; revised 9 March 2018; accepted 20 March 2018; published online 27 April 2018)

In order to convey pragmatic functions, a speaker has to select an intonation contour (the tune) in addition to the words that are to be spoken (the text). The tune and text are assumed to be independent of each other, such that any one intonation contour can be produced on different phrases, regardless of the number and nature of the segments they are made up of. However, if the segmental string is too short, certain tunes—especially those with a rising component—call for adjustments to the text. In Italian, for instance, loan words such as “chat” can be produced with a word final schwa when this word occurs at the end of a question. This paper investigates this word final schwa in the Bari variety in a number of different intonation contours. Although its presence and duration is to some extent dependent on idiosyncratic properties of speakers and words, schwa is largely conditioned by intonation. Schwa cannot thus be considered a mere phonetic artefact, since it is relevant for phonology, in that it facilitates the production of communicatively relevant intonation contours. © 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5030923>

[BVT]

Pages: 2474–2486

## I. INTRODUCTION

Common to all approaches to intonation is the assumption that the intonation contour is independent of the words that bear it. Not only is intonation independent in terms of the meaning it conveys: “Intonation operates in its own sphere” (Bolinger, 1957), but it is also independent of the length of words and their segmental makeup: “A pattern of speech melody in intonation is independent of words” (Abercrombie, 1967). Autosegmental approaches to intonation make this independence explicit in that the intonation contour, made up of tones—the tune—is on a separate tier from the words, syllables, and segments—the text (Lieberman, 1975; Leben, 1976; Goldsmith, 1976; Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988; Ladd, 2008).

For it to be perceived, any tune needs to occur on segmental material of high intensity and rich harmonic structure. Consequently, the independence of tune and text might be compromised if there is insufficient or inadequate segmental material available for the realisation of the tune. Commonly, intonation involves a sparse distribution of tones, such that the number of tones is outnumbered by the number of tone-bearing-units in the text (commonly syllables). However, the tones are not spread out evenly over the utterance, but are instead associated at strategic privileged

positions: heads (e.g., heads of feet or prosodic words, namely, syllables with stress) and edges of constituents (e.g., intonation phrases). If these positions are close together (such as when a stressed syllable is final in a phrase), it can lead to crowding of tones onto one syllable. The text may then be inadequate for bearing the tune, especially if the tune is complex and the syllable is short or contains voiceless segments. In such cases, adjustments can be made to either the tune or the text. The nature of these adjustments depends on syntagmatic, paradigmatic, and language-specific factors (see Hanssen, 2017, or Roettger, 2017, for recent overviews).

### A. Adjustments to the tune

If the segmental tier offers too little tone bearing material for the realisation of a tonal sequence, the pitch contour can be modified. The first studies reporting such modifications were not on intonation, but on lexical accent. In Swedish, both accent 1 and accent 2 words have a falling pitch contour, represented as a sequence of high and low tones. The difference between minimal pairs with this lexical tone distinction lies in the alignment of the high tone: earlier for accent 1 and later for accent 2 (Bruce, 1977). In their seminal work, Erikson and Alstermark (1972) discuss how the realisation of a lexical pitch accent is adjusted as a function of the segmental structure. On the one hand, they observed that the pitch movement is often reduced with decreasing vowel length, i.e., the pitch movement is under-shot, with the fall after the high tone simply ending before it

<sup>a)</sup>Electronic mail: [martine.grice@uni-koeln.de](mailto:martine.grice@uni-koeln.de)

<sup>b)</sup>Also at: Northwestern University, Department of Linguistics, 2016 Sheridan Rd., Evanston, IL 60208, USA.

reaches what is assumed to be its low target. Moreover, if the voiced material is particularly short, such as when the syllable contains a short vowel and a voiceless coda consonant, the fall in pitch may be entirely missing. This mechanism has been called “truncation.” In addition to truncation, Erikson and Alstermark (1972) also discuss “rate adjustments,” where the pitch movement is realised more rapidly on shorter vowels. This mechanism was subsequently named “compression” (Bannert and Bredvad-Jensen, 1975).

Both truncation and compression have since been attested in intonation. Grønnum (1989) found truncation in Danish rise-falls and in Northern German falls. Grice (1995) found truncation in Palermo Italian, with complex rising-falling contours losing the final fall altogether, even if the syllable is fully voiced. Both truncation and compression were also found in the standard varieties of German and English by Grabe (1998), who systematically manipulated the segmental material available to realise the intonation contour. This manipulation ranged from a disyllabic word with a long vowel (*Sheafer* /ʃi:ʔə/, *Schiefer* /ʃi:ʔə/) through a monosyllabic word with a long vowel (*Sheaf*, *Schief* /ʃi:f/) to a monosyllabic word with a short vowel (*Shift*, *Schiff* /ʃɪft, ʃɪf/). The stimuli contained voiceless consonants, in this way restricting the phonetic opportunity for realising tonal movements to the vowel only. Her results indicate that Southern Standard British English compresses both falls and rises, whereas Northern Standard German truncates falls and compresses rises. Truncation was also found in high fall-rise contours in German and Dutch (Lickley *et al.*, 2005; Ladd, 2008). Here the fall is missing, resulting in a high rise. Rathcke (2009) confirmed truncation on rising-falling contours in German, but found that falling contours can be partially compressed. Moreover, truncation and compression have been found to be dialect dependent (Bannert and Bredvad-Jensen, 1975; Grabe *et al.*, 2000; Peters, 2006; Roseano *et al.*, 2015), often involving mixed strategies, and have more recently been found to be speaker specific too (Prieto and Ortega-Llebaría, 2009).

A different strategy to deal with problems in tune-text-association is a temporal shift of the tonal target. If there is not enough segmental material to realise a sequence of tonal targets, one or multiple tonal targets may be realised earlier, or anticipated. A case in point is Neapolitan Italian, in which questions and statements with final narrow focus are expressed with similar tonal events, i.e., a rise-fall in pitch phrase finally. Even though these contours appear to be very similar in certain contexts, they have been shown to differ in the alignment of the high tone, which reaches its target later in the accented vowel in questions than in contrastive statements (D’Imperio and House, 1997). Questions in Italian, however, can also have an additional final rise making the tonal movement more complex (a rise-fall-rise as opposed to rise-fall). The presence vs absence of an extra final rise is dependent on dialect and speaking style (Savino, 2012). Cangemi and Grice (2016) looked at the alignment of the high target of the pitch accent across instances with and without an utterance-final rise. They found strong evidence for the high target shifting to the left when a phrase-final rise is present, even to the point of being earlier than the

statement target. It is argued that the tonal sequence starts earlier to ensure the realisation of all tonal targets in this complex contour. The tendency of tones to be anticipated in tonal crowding contexts has been observed in many languages (Steele, 1986; Caspers and van Heuven, 1993; Prieto *et al.*, 1995; D’Imperio and House, 1997; Prieto *et al.*, 2005; Schepman *et al.*, 2006; Mücke *et al.*, 2009; Grice *et al.*, 2015; Hanssen, 2017; Roettger, 2017).

## B. Adjustments to the text

Tune adjustments, such as truncation, compression, and anticipation of the pitch movement, are not the only ways to resolve a mismatch in tune and text. Adjustments can also be made to the text. This can be achieved by lengthening the segmental material. This type of adjustment has been found in a number of languages, typically when there is a complex intonation contour, involving at least one change of direction, and when this contour is either on a monosyllable or on a word with final stress, leading to crowding of the pitch accent and boundary tones onto one syllable.<sup>1</sup>

Lengthening of segments has been observed for German, both for Standard German (Gartenberg and Panzlaff-Reuter, 1991) and the Hamburg variety (Gilles, 2005). Gilles (2005) found that monosyllables are longer with a fall-rise intonation than with falling contours. Similar lengthening was found in Spanish and Catalan by Prieto and Ortega-Llebaria (2009), also with a complex contour. They found that, in words with final stress, syllables are longer in rise-falls than in simple falls. Similar results have been reported for Bari Italian yes-no questions, which are typically realised (in read speech) with an accentual rise followed by a fall-rise. If a phrase-final accented syllable bears the rise-fall-rise, it is considerably lengthened, as compared to the same syllable in neutral statements, in which there is a simple fall (Grice *et al.*, 1997; Refice *et al.*, 1997). Frota (2002) also reports on lengthening of phrase final vowels in fall-rises. Heston (2014) reports on lengthening of utterance-final vowels in Fataluku when they are accompanied by a final rise-fall. In all these cases there is lengthening when the phrase-final syllable is accented (in the case of Fataluku phrase accented) and the contour is complex. This means that it may not be tonal crowding *per se*, but rather the necessity to realise complex contours that appears to be driving the lengthening.

In addition to lengthening existing segments, some languages facilitate tonal realisation by inserting additional segments. In Tashlhiyt, a language that has long consonantal sequences, the insertion of a schwa-like element has been reported to correlate with tune requirements in intonation (Grice *et al.*, 2015; Roettger, 2017), just as in setting text to music in this language, where schwa can be on a separate note (Dell and Elmedlaoui, 2013). While these elements have been argued to play no role in syllabification, it could be argued that schwa in Tashlhiyt is a structural element that enables the realisation of functionally relevant tonal movements. Important for the present study is that this schwa can be inserted phrase finally after the final lexical consonant. Frota (2002, 2014) and Frota *et al.* (2016) report on similar insertions of non-lexical vowels after sonorants in the

standard variety of European Portuguese: A high central vowel is inserted after the phrase-final lexical syllable—again, as in the lengthening cases above—if there is a complex tonal movement, in this case a fall-rise. Hellmuth (2018) also reports the insertion of schwa in Tunisian Arabic, again in conjunction with a complex tonal movement, in this case a rise-fall, in the intonation contour used in questions. Importantly, in Tashlhiyt, Standard European Portuguese, and Tunisian Arabic, a vowel is only inserted some of the time, *even if* the conditions mentioned (a complex tonal movement) are met. Thus, it is clear that in these three languages other factors play a role in determining text adjustments as well as intonation. We return to these languages when interpreting our results.

The above findings suggest that tune and text are less independent when intonation contours are complex, in particular when the necessity for tonal realisation conflicts with the extent to which segments lend themselves to a clear manifestation of these tones. A possible solution to this functional dilemma is inserting new segmental material that facilitates the realisation of the tune.

### C. Adjustments to the text in Italian loan words

A promising area for exploring the nature of inserted segments is the pronunciation of loan words in Italian. Recall that tune-related durational adjustments have already been reported for one variety of this language in words ending in a vowel (Grice *et al.*, 1997; Refice *et al.*, 1997). In the Bari variety, the final vowel in words such as bambù (/bam'bu/ “bamboo”) is considerably lengthened in yes-no questions, which have a complex rise-fall-rise intonation contour. Since the words investigated have stress on the final syllable, which bears the nuclear pitch accent, the requirement to produce the intonation is assumed to lead to this lengthening. If intonation can condition adjustments to the duration of lexical vowels, it is also possible that it can play a role in conditioning the insertion of non-lexical vowels. Such non-lexical vowels are sometimes appended to words ending in a consonant, such as bus (/bus/) and bar (/bar/). This insertion of vowels is reportedly common in Central and Southern varieties of Italian (Lepschy and Lepschy, 1992; Bertinetto, 1985), and has even been the subject of debate as to whether these vowels should be represented in the orthography.

What happens on consonant-final words might be expected to be of scant relevance for Italian phonology, given that the language has a small number of such words in its native vocabulary. However, since Italian has incorporated a great number of consonant-final words in recent years, including many proper nouns (D’Achille, 2010), the question as to how they are pronounced is highly topical. In fact, the incorporation of these loan words has been argued to represent “the main phonological innovation” of the 20th century (Klajn, 1972; cited in Repetti, 1993). The borrowing process is active, and includes words from music, technology, computing, business, media, politics, and law (D’Achille, 2010), such as *chat*, *blog*, *staff*, *web*, *stop*, *link*, and so on.

The pronunciation of these words is reportedly variable, both in terms of the properties of the consonant itself—it can have varying durations—as well as in the presence or absence of a non-lexical vowel (Bertinetto, 1985; Krämer, 2009; Repetti, 2012; Broniś, 2016) and even in terms of the phonetic properties of this vowel. One strong argument for the phonological status of the non-lexical vowel is that it can be accompanied by a lengthened (geminated) consonant, suggesting that the insertion of schwa is structure-building, forming a separate syllable with the second half of the geminate as its onset, e.g., /'bus.sə/. However, the phonetic properties of both consonant and vowel appear to be prone to inter- and intra-speaker variability, as evidenced in the transcriptions [stɔp̚] [stɔp̚ː] [stɔp̚ː], as well as [stɔp̚ə] and, for some varieties also [stɔp̚e] (Bertinetto, 1985; Repetti 2012). These observations indicate that the insertion of schwa (and any lengthening of the consonant preceding it) is somewhat gradual in nature. In these cases, it remains an open question, which role these inserted elements play in the phonology of a language, i.e., whether they are to be treated as segments in their own right, possibly contributing to syllable structure, or simply as transitional elements with no phonological status. In fact, the variability—both in terms of the presence or absence of schwa, as well as in terms of its phonetic properties—might be taken as evidence for a less entrenched phonological status.

Such considerations can be related to the typology of inserted vowels proposed by Hall (2006). She has argued that there are two distinct kinds of inserted vowels (see also Harms, 1976; Levin, 1987; Warner *et al.*, 2001; Silverman, 2011). On the one hand, there are epenthetic vowels, i.e., segments that have stable forms and distributions and are visible to the phonological system. On the other hand, there are intrusive (or excrescent) vocoids, i.e., phonetic artefacts that do not have any phonological status and are thus invisible to the phonological system. The literature on Italian loan words appears to provide evidence in both directions.

The variety of Bari is well suited to a study of schwa insertion in Italian, since it has a highly complex tune in polar questions: rise-fall-rise, or rise-fall, as opposed to certain other varieties of Italian, in which these questions have simple rises (Savino, 2012; Gili Fivela *et al.*, 2015). Moreover, Bari Italian has two further rising tunes, occurring in prefinal position in lists (high rise) and in earlier positions, referred to as non-final (low rise), discussed in more detail in Sec. ID below.

As mentioned above, yes-no questions in this variety have already been shown to incur an increase in the duration of stressed final lexical vowels (e.g., in *bambù*) in this variety. Moreover, being in the South, Bari Italian is expected to make liberal use of schwa insertion after a word final consonant (Lepschy and Lepschy, 1992; Bertinetto, 1985). As in other varieties of Italian, the functional load of intonation is high, given that intonation is the sole cue used for distinguishing questions from statements.

It is important to remember here that consonant final loan words are not always monosyllabic, as might be suggested by the examples commonly cited (and provided in the paragraphs above). If they are polysyllabic—words like

*server, tunnel, twitter, network*—the final syllable is rarely stressed. Although word final schwa insertion in polysyllabic words has received little attention, to date, it has been attested in the Roman variety (Broniś, 2016). Since these words do not have final stress, the pressure to adjust the text should be weaker, there being two syllables to realise the pitch accent and following boundary tones, resulting in less tonal crowding.

The fact that consonant-final words may be subject to differing degrees of tonal crowding leads to the question as to whether reported variation in the phonetic properties of the inserted vowel might be related to variation in the pressure to adjust the text.

#### D. Research question and hypotheses

The main research question is whether in Bari Italian the requirement to produce a particular tune conditions the insertion of a non-lexical vowel as well as conditioning the duration of this vowel. Tunes involving a rise, henceforth rises, (rise-fall-rise L+H\* L-H%, low rise L\* L-H%, and high rise H\* H-H%) will be compared to tunes without a rise (low fall H+L\* L-L%), henceforth falls (see Fig. 1 for schematic contour shapes).

Research so far has mainly been concerned with segmental adjustments in yes-no questions. Across the languages in which adjustments have been reported, this illocutionary act has a wide range of different intonation patterns (although all contain a rising component), as compared to neutral statements, which generally have a low falling intonation pattern. To tease apart whether adjustments are affected by questions *per se* or by the intonation pattern, we also look at other contexts in which there are intonation patterns that differ from the neutral declarative one. Specifically, Bari Italian is known to use distinct intonation patterns on items in lists, depending on which position they occupy (Savino, 2001, 2004; Savino *et al.*, 2006). Apart from the final item, which has a low fall as in neutral statements, other list items have rises: a low rise in non-final position and high rise in prefinal position.

The majority of studies investigating segmental adjustments involve the crowding of tones on the final stressed syllable of a word. However, since at least one variety of Italian reportedly inserts schwa in words with penultimate stress, it is necessary to investigate the effect of different tunes on such words too. Disyllables with a trochaic structure allow more time for tones associated with the accented syllable to

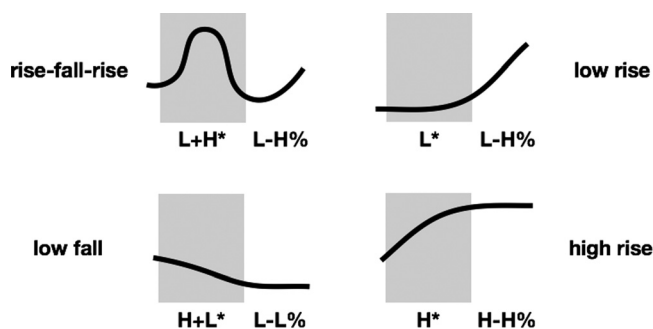


FIG. 1. Stylised pitch contours for rise-fall-rise, low fall, low rise, and high-rise contours, with the shaded area representing a stressed syllable.

unfold, making it difficult to argue that yet another syllable is necessary for a tune to be adequately realised.

The present study attempts to reject the following null hypothesis: Tune-driven text adjustments do not condition schwa insertion in Bari Italian, i.e., the frequency of occurrence of schwa—and, if present, its duration—are independent of both the tune and the metrical structure in the target word. Thus, we test the following alternative hypotheses:

H1a: Rising tunes lead to more frequent occurrence of schwa than falling tunes.

H1b: Rising tunes lead to longer schwas (when present) than falling tunes.

H2a: Schwa is inserted more frequently on monosyllables than trochaic disyllables.

H2b: Schwa (when present) is longer on monosyllables than trochaic disyllables.

## II. METHOD

### A. Speech material

To test these hypotheses, we elicited questions and statements (rise-fall-rise L+H\* L-H% and low fall H+L\* L-L%, respectively) and lists (low rise L\* L-H%, high rise H\* H-H%, and low fall H+L\* L-L% for non-final, prefinal, and final positions, respectively) with the nuclear pitch accent on proper nouns (person names), a common source of loans in the language. Target words were either monosyllables or trochaic disyllables, leading to pitch accent placement on either the final or the penultimate syllable of the phrase, respectively. In the former, the pitch accent and boundary tones crowd together on one syllable, in the latter they can potentially spread over two syllables.

Target words consisted of ten monosyllabic and six disyllabic names (Bill, Moll, Tim, Dan, Dag, Fred, Chris, Jeff, Matt, Dick for the monosyllables, and Caleb, Colin, Carol, Edith, Derek, Dennis for the disyllables).<sup>2</sup> These target words were elicited in five prosodic conditions: (polar) questions and (neutral) statements, and in three distinct positions in lists. Each list consisted of six names, and names were coded as non-final (NF), prefinal (PF), or final (F). For each prosodic condition, target words were produced within an appropriate context, as follows (see also Fig. 2):

- (1) Question (with following answer as context):

Ha chiamato [target name]? ‘Did [target name] call?’

No, ha chiamato [name]. ‘No, [name] called.’

e.g., Ha chiamato Jeff?

Each *target name* was produced in the same carrier sentence. The answer was always negative as above, and the *name* in the answer was varied so as to introduce variety into the task.

- (2) Statement (answer to contextualising question):

Chi ha chiamato? ‘Who called?’

Ha chiamato [target name]. ‘[target name] called.’

e.g., Ha chiamato Jeff.

For this condition, the contextualising question and carrier phrase for the answer was constant across all target words.

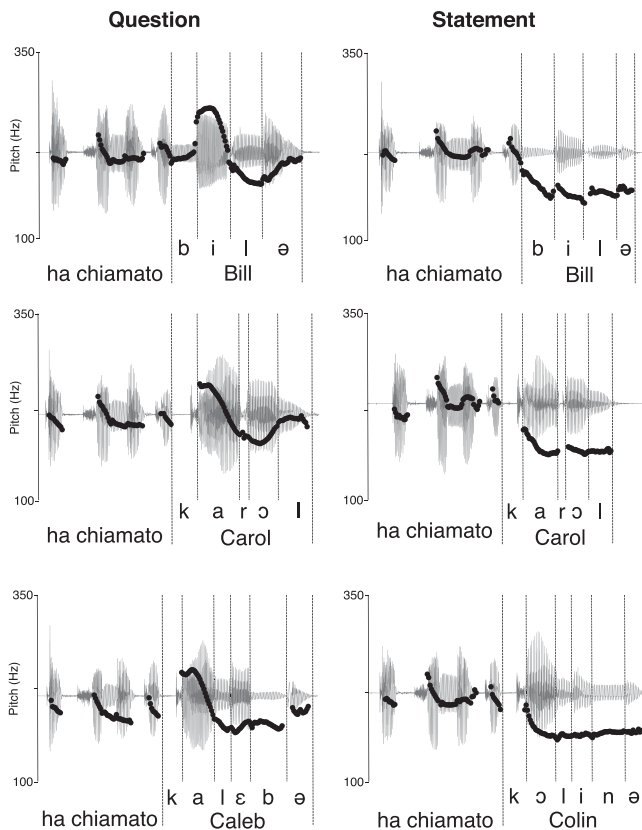


FIG. 2. Representative waveform and  $F_0$  contour for questions (left column) and statements (right column) for monosyllabic target words with schwa (top row), disyllabic target words without schwa (middle row), and disyllabic target words with schwa (bottom row). All examples are produced by the same speaker.

(3) NF, PF, and F positions in a list of names:

Ecco la lista dei nomi: 'Here is the list of names:'

[NF target], [NF target], [NF target], [NF target], [PF target], [F target].

e.g., Dan, Colin, Dennis, Moll, Matt, Fred

Three sets of lists were constructed: (1) lists with non-final target names (NF list)—this target can be in one of the first four positions, (2) lists with prefinal target names (PF list), and (3) lists with final target names (F list). Thus, only one name was treated as a target in each list.

In sum, there were 160 items in total (16 target words  $\times$  5 prosodic conditions  $\times$  2 repetitions) per speaker.

## B. Participants and procedure

Ten native Bari Italian speakers participated in the recording session on a voluntary basis. They were all female (aged 22–29 years) and undergraduate students of psychology at the University of Bari.

Speakers were seated in front of a computer screen, wearing a headset microphone (AKG C520, Vienna, Austria) connected to a Marantz PMD 661 digital recorder (Kanagawa, Japan). Each target phrase was presented on the screen along with its context (via a PowerPoint presentation, Microsoft, Redmond, WA). Speakers were instructed to read the whole of the text on the screen first silently and then aloud at a normal pace and in a natural way. No word was

highlighted, and speakers were not told which parts were context and which were target phrases and words. If a speaker was unsatisfied with their production, either because they felt it was unnaturally produced or because there was a dysfluency, they were allowed to repeat the whole text on the screen (i.e., including the context). In this case, the repetition was taken for analysis instead of the first production. Speakers were also allowed to take a break any time they needed, which was at least once every 20 stimuli.

Stimuli in context were presented in five separate blocks, according to prosodic condition (question, statement, NF list, PF list, and F list) and target words were randomised in each block. The order of presentation of blocks was randomised for each speaker. There were no fillers.

## C. Acoustic analysis

Target words were manually segmented and annotated with Praat (Boersma, 2001) employing the following labeling criteria: We identified segment boundaries in the target word in the acoustic waveform. To do this, we displayed an oscillogram and a wide-band spectrogram simultaneously. All segmental boundaries of vowels and consonant were labeled at abrupt changes in the spectra at the time at which the closure was formed or released: this was the case for the nasals, the laterals (especially in the spectra for the intensity of higher formants), and the fricatives (at random noise patterns in the higher frequency regions).

The labeling of potential schwa was not always straightforward. We thus adopted a liberal approach, labeling as a schwa any interval presenting periodic vibrations accompanied either by a local increase in the signal energy at the consonantal release, and/or any interval after the consonantal release with formant structure or energy in the  $F_2/F_3$  region characteristic of vowels. In some cases, deciding on the presence of these acoustic features was very difficult. Thus, we kept track of these ambiguous cases and ran all statistical analyses with and without these problematic cases. Exclusion of these cases did not make any difference.

## III. RESULTS

### A. Intonation contours

Although speakers were not explicitly instructed to produce a particular intonation contour, they were consistent in their productions, both within and across speakers. The intonation contours in both data sets corresponded to our expectations from previous studies, which were also based on read speech. In the question-statement dataset, questions were produced predominantly with a rise-fall-rise ( $L+H^* L-H\%$ ) and occasionally with a rise-fall ( $L+H^* L-L\%$ ), whereas statements had a low fall ( $H+L^* L-L\%$ ). See Figs. 2 and 3 for examples. It is evident from the examples in Fig. 2 that the complex rise-fall-rise contour in questions takes up the entirety of the segmental material ( $/'bilə/, /'karəl/, /'kaləbə/$ ), the final rise being on the schwa in Bill and Caleb, or on the final syllable of Carol. By contrast, in statements the full extent of the fall is achieved by the middle of the stressed

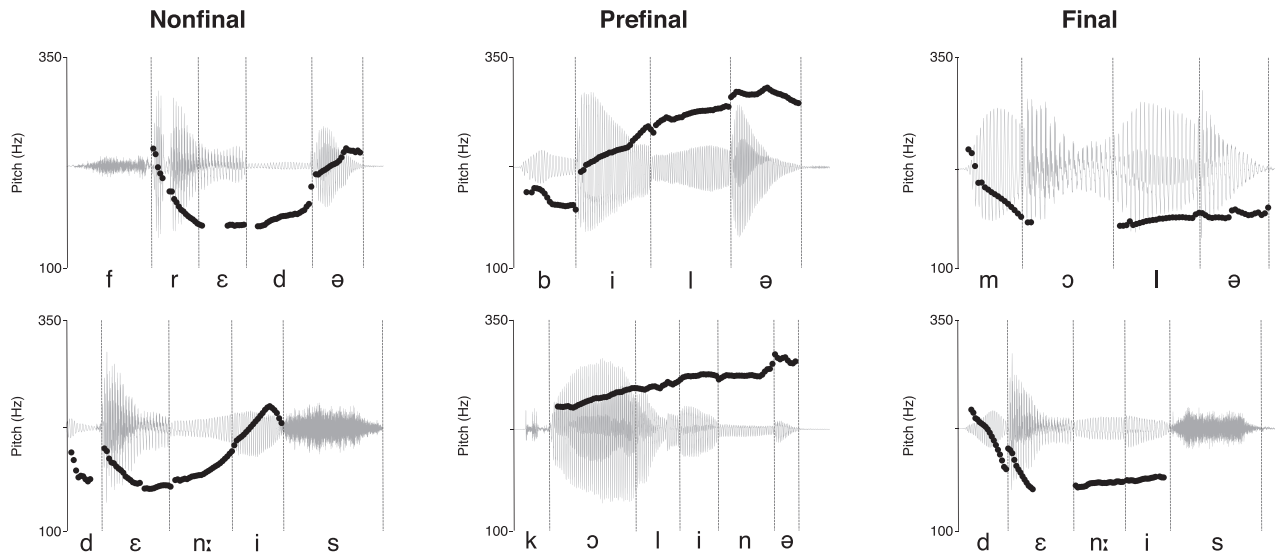


FIG. 3. Representative waveform and  $F_0$  contour for selected items representing different positions in a list (non-final, prefinal, and final) for monosyllabic target words with schwa (top row) and disyllabic target words (bottom row). All examples are produced by the same speaker (see footnote 7).

syllable (in /'bilə/, /'karəl/, and /'kəlinə/). After that there is a low flat stretch of pitch.

In the lists, NF items were produced with a low rise (L\* L-H%), PF items with a high rise (H\* H-H%), and F items with a low fall (H+L\* L-L%; see Fig. 3 for examples). In the NF examples, the rise occurs mainly after the stressed syllable, whereas in the PF examples, the pitch begins higher and rises throughout the whole word, including the stressed syllable. The F items in lists have the same intonation pattern as the statements in the question-statement dataset.

## B. Inferential analysis according to research hypotheses

First, we tested our research hypotheses, i.e., we attempted to reject the null hypothesis that the frequency of occurrence of schwa and, if present, its duration is independent of both the tune and the metrical structure in the target word.

All data were analysed and plotted using R (R Core Team, 2015) and the packages *afex* (Singman et al., 2017), *ggeffects* (Lüdtke, 2017), *lme4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017), and *tidyverse* (Wickham, 2017). To analyse categorical data, mixed logit models with a binomial error function were fitted to the binomial measurement of whether a schwa was present or not. To analyse continuous dependent variables, mixed linear regression models were fitted to schwa duration. We performed analyses on two separate subsets of the data: We analysed all data elicited by the question-answer materials (640 data points) and all data elicited by the list materials (960 data points) separately.

The critical predictors were sum-to-zero contrast-coded TUNE (questions vs statements in the question-answer subset and NF, PF, and F in the list subset), sum-to-zero contrast-coded METRICAL STRUCTURE (monosyllabic vs disyllabic), and their interaction. The random effects components varied between models. They are specified for each model in Sec. III,

Results. Since generalised linear mixed effect models are well-known to fail to converge on parameter estimates, especially with logistic regressions, some of our models are not specified for by-word and by-speaker random slopes. Our model selection process unfolded as follows: We started with the maximal random effect structure, including the by-speaker slope for the interaction of TUNE and METRICAL STRUCTURE and the by-word slope for the TUNE. If the model (and respective reduced models) did not converge, we subsequently reduced the random slope complexity until we reached the maximally converging model (see accompanying R scripts for the selection process).

We calculated  $p$ -values based on likelihood ratio test. They are obtained by comparing a model in which the tested effect and all higher order effects (e.g., all two-way interactions for testing a main effect) are excluded with a model in which only effects up to the order of the tested effect are present and all higher order effects absent. In other words, there are multiple full models, one for each order of effects. Consequently, the results for lower order effects are identical of whether or not higher order effects are part of the model or not. In line with standards of reproducible research (Peng, 2011), the data table and the scripts for the statistical analyses are made available and can be retrieved.<sup>3</sup>

### 1. Predicting the presence of schwa

In general, there were many instances of schwa throughout both of the data sets, with schwa being present in 79% of all target words in the question-statement data set and 74% of all target words in the list data set.

To answer the question as to whether tune and metrical structure (monosyllabic vs disyllabic) affects the presence of schwa (H1a and H2a), we fitted mixed logit models to schwa presence for the question-answer subset and the list subset separately. We included random intercept for words and by-speaker random slopes for METRICAL STRUCTURE for the question-answer subset. We included random intercepts for both words and speakers in the list subset.<sup>4</sup>

For the question-statement subset, there were significant effects of TUNE ( $p = 0.002$ ), METRICAL STRUCTURE ( $p = 0.003$ ), and their interaction ( $p < 0.0001$ ), such that statements exhibit fewer schwas than questions, disyllables exhibit fewer schwas than monosyllables, and the decrease in number of schwas for disyllables is stronger in statements than in questions [see Fig. 4(a)]. Similarly, for the list subset, there were significant effects of TUNE ( $p < 0.0001$ ), METRICAL STRUCTURE ( $p < 0.0001$ ), and their interaction ( $p < 0.0001$ ), such that monosyllables exhibited more schwas than disyllables. Moreover, the increase in the probability of schwa occurrence in monosyllables (as opposed to disyllables) is strongest for words in NF position, followed by PF and F words [see Fig. 4(b) and Table I].

## 2. Predicting the duration of schwa

We addressed the question as to whether the TUNE and METRICAL STRUCTURE in the target word affected not only the presence but also the duration of schwa. To do this, we fitted mixed linear regression models to schwa duration in all instances exhibiting a schwa for the question-statement and list data sets separately. We included random intercepts for words and speakers. Additionally, we included by-word random slopes for the factor TUNE and by-speaker random slopes for the interaction term of TUNE and METRICAL STRUCTURE.

For the question-statement subset, there were significant effects of TUNE ( $p = 0.007$ ), METRICAL STRUCTURE ( $p = 0.0001$ ), and their interaction ( $p = 0.002$ ), such that schwas in statements exhibited smaller durations than in questions, schwas

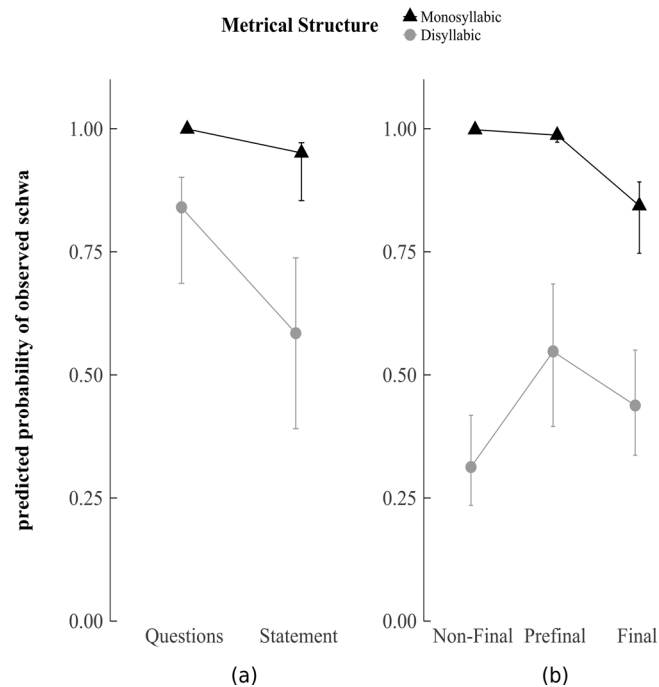


FIG. 4. Predicted probability of schwa occurrences as a function of the tune (x axis) and metrical structure in the question-answer subset (a) and list subset (b), respectively. Error bars indicate  $\pm 1$  standard errors (SEs) from the mean, taken from the model described above. Note that SEs are based on logit calculations and naturally decrease in the probability parameter space approaching the boundaries 0 and 1. Consequently, the standard error of certain estimates approaches zero and is visually undetectable.

TABLE I. Measured proportion of observed schwa and, when present, its duration as a function of tune and metrical structure in the target word.

	Monosyllabic		Disyllabic	
	Proportion (%)	Duration (ms)	Proportion (%)	Duration (ms)
Question	99	121	70	90
Rise-fall-rise				
Statement	80	84	53	76
Low-fall				
Non-final	100	107	36	62
Low-fall-rise				
Prefinal	97	103	53	61
High rise				
Final	78	87	45	80
Low Fall				

in disyllables exhibited smaller durations than in monosyllables, and the increase of schwa duration for monosyllables is stronger in questions than in statements [see Fig. 5(a) and Table I]. Similarly, for the list subset, there were significant effects of METRICAL STRUCTURE ( $p < 0.0001$ ), and its interaction with TUNE ( $p = 0.0004$ ), such that schwas in disyllables exhibited smaller durations than in monosyllables and the difference between monosyllables and disyllables was conditional on its tune: The schwa duration difference between monosyllables and disyllables was smaller for F words compared to NF and F words [see Fig. 5(b)]. TUNE had no independent main effect on schwa duration ( $p = 0.53$ ).

In sum, the data provide evidence against the null hypothesis and in favour of the alternative hypotheses (H1-2). The above results suggest effects of the tune and the metrical structure on both the presence of schwa and its

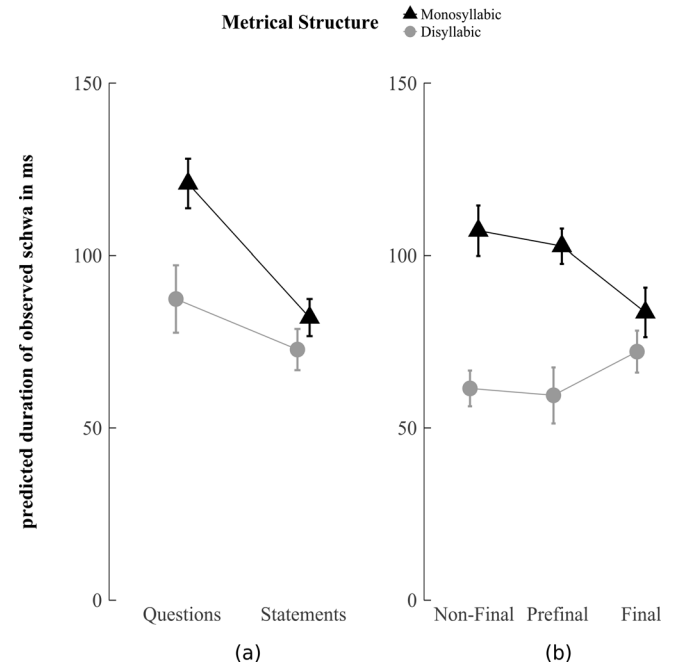


FIG. 5. Predicted duration of schwa as a function of tune (x axis) and metrical structure in the question-answer subset (a) and list subset (b), respectively. Error bars indicate  $\pm 1$  SEs from the mean. SEs are taken from the model described above.



duration. In the question-statement data set, questions, characterised by a rise-fall-rise, are more likely to exhibit a schwa and if schwa is present it is longer compared to statements, characterised by a very small pitch movement (a low fall). While monosyllabic words surfaced with schwa in the majority of cases, disyllables surfaced less often with schwa. This asymmetry is more pronounced for statements. Regarding the duration of schwa, there is a smaller effect of metrical structure in statements than in questions.

These patterns are mirrored in the list data set. Words in PF position, characterised by a high rise, are more likely to exhibit schwa and if schwa is present it is longer, compared to words in list-final position, characterised by a fall. Again, monosyllabic words surfaced with schwa in the majority of cases, disyllables surfaced less often with schwa, and schwa had a longer duration in contexts with rising pitch movements (NF and PF positions) as opposed to falling ones (F position).

Across the two datasets, the statements and F words in lists have similar schwa durations. This is unsurprising, since they are both in final position and have the same tune (low fall).

Our results are very much in line with our formulated hypotheses and the assumption that the necessity to realise tonal movements affects the realisation of schwa. If the word is monosyllabic, the text is suboptimal for bearing a pitch movement. The presence of a schwa in such cases enables the tune to be realised on more voiced material. The presence of schwa is further affected by the tonal movement to be realised. A more complex tune (rise-fall-rise) needs more space to be realised than a simple tune (fall), thus schwa is more likely to be present in questions than in statements, and if it is present, it is longer. In fact, for monosyllabic words realised in questions, all productions but one exhibited a schwa. Moreover, in the list dataset, NF and F monosyllables almost all had schwa, showing that, in the monosyllabic condition, rising tunes were more likely to be produced with schwa than the low falling tunes. This was not always the case in disyllables in this dataset: Here the low rise (non-final position in list) led to *fewer* schwas than the low fall (final position in list).

When looking at these results, it is important to note that schwa generally surfaces very frequently in our corpus, with its presence and duration characterised by a great deal of variability beyond the hypothesised impact of tune and metrical structure. In Sec. III C, we are concerned with the question as to how far other factors can account for this variability.

### C. Explorative analysis of schwa presence and duration

To further explore which other factors contribute to the presence and duration of schwa, random forests analyses were applied (Breiman, 2001), implemented by the *party* package (Strobl *et al.*, 2008). Random forests analysis is a data mining technique used for classification and has already been applied to several phonetic data sets (e.g., Tagliamonte and Baayen, 2012; Winter and Grawunder, 2012; and in a more closely related study by Roettger, 2017, on vowel insertion in Tashlhiyt). It is a so-called “ensemble method.” A multitude

of decision trees is constructed (500 in this case). Each tree takes a set of variables and sees which variable best splits the data according to a particular criterion. Each tree is built on a random subset of variables and data. The final classification is based on the overall ensemble of trees. Random forests allow us to explore which factors are independently relevant for determining the presence vs absence of schwa or its duration, respectively, i.e., although factors might correlate with each other, this ensemble method leads to an estimate of each individual factor contribution independently of the other factors. The following factors were included in the analysis: Factors capturing idiosyncratic properties of speaker and the word accounting for inter-speaker and word-specific variability. Next, we included the identity of the word final consonant, which appears to be relevant for non-phonological vowels in other languages (e.g., Ridouane and Fougeron, 2011; Frota, 2002, Frota *et al.*, 2016; Hellmuth, 2018; Kwon, 2017): Factors capturing consonants were categorically coded as phonologically  $\pm$  voiced,  $\pm$  sonorant, and  $\pm$  fricative. We added a factor controlling for word-level durational adjustments looking at the duration of a reference vowel (the stressed vowel) in milliseconds (ms) (see, e.g., Kilbourn-Ceron and Sonderegger, 2018, for vowel devoicing). Finally, we included the two factors from our confirmatory analysis: A factor capturing metrical characteristics of the target word coded as metrical structure (monosyllables vs disyllabic trochees); and a factor capturing prosodic characteristics of the contour coded as tune (question vs statement in the question-statement data set, and the position in the list: non-final, prefinal, and final in the list data set).

Figure 6 ranks predictors according to their relative importance to predict the dependent variables. There is no threshold as to what is important enough (in the traditional sense of important enough to reject the null hypothesis). Only by comparing the predictors' relative contribution can we generate new hypotheses about relevant relationships.

In Fig. 6, it is apparent that a number of different factors are important for predicting whether speakers produce a schwa or not [Fig. 6(a)] and if so, how salient this schwa is acoustically in terms of its duration [Fig. 6(b)]. As expected, the analysis of the presence and duration of schwa reveals a large impact of the tune, which is unsurprising, given our confirmed hypotheses discussed in Sec. III B. The effect of metrical structure turns out to be comparatively weak for schwa presence and negligible for schwa duration. The latter finding is surprising and suggests that a certain amount of variance might be explained by other factors such as idiosyncratic properties of the words in our corpus.

For both analyses, the idiosyncratic factors speaker and word are highly ranked. This ranking reflects the high inter- and intra-speaker variability that is reportedly a common characteristic of the production of loan words in the language. Similar strong effects of idiosyncratic properties of speakers and words have been reported for schwa insertion in Tashlhiyt (Roettger, 2017), in which it has been suggested that schwa insertion is at least partly dependent on gender (more schwa by women) and place of upbringing (more schwa by speakers coming from urban areas). Although our speaker sample is relatively homogenous (all women, all

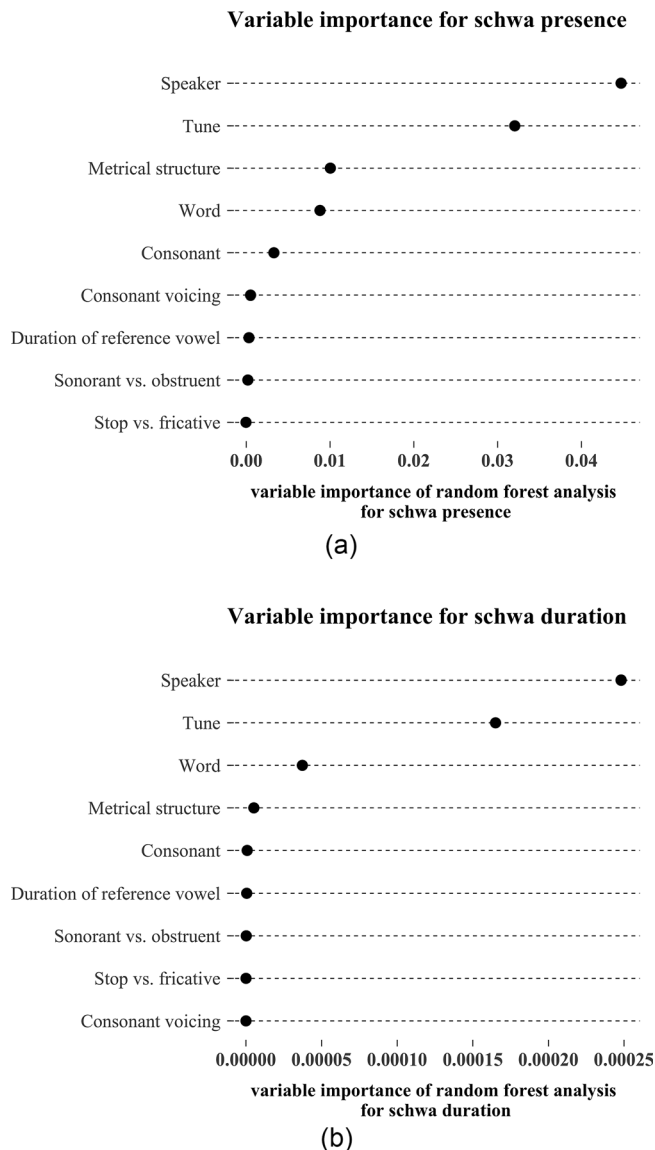


FIG. 6. Variable importance measure generated by random forests predicting presence of schwa (a) and schwa duration (b). Note that the units of variable importance are non-informative beyond capturing the relative contribution of each factor compared with the others.

students of psychology, all from the same geographical area), there remains a substantial degree of variability across individuals. It is important to emphasise here that the tune explains almost as much variability as the speaker variability, indicating a strong impact of the tune-text requirements on the presence and duration of schwa.

Contrary to expectations, phonetic properties of the word final consonant do not explain residual variation, despite factors grouping consonants into phonetically motivated classes (voiced vs voiceless; stop vs fricative; sonorant vs obstruent). Moreover, the analyses do not show any relevance of the duration of the reference vowel, neither in the presence nor in the durational properties of schwa.

#### IV. GENERAL DISCUSSION

We have shown that the insertion and acoustic prominence of schwa in Bari Italian is related both to the tune and

to the metrical structure of the target word on which it is realised. Whilst schwa occurred almost all of the time on monosyllables with rising intonation contours (including contours with a rising component), it occurred less frequently on monosyllables with falling contours. Furthermore, there were more schwas on monosyllables in general than on disyllables. Within the disyllables, the effect of rising intonation was less clear-cut than for monosyllables, in that schwa was very frequent on two of the rising contours (rise-fall-rise in questions and high rise in non-final list items), but was infrequent on the low-rise tune (found on non-final list items).

Our inferential analyses indicate that the insertion of schwa can be seen as an adjustment of the text in response to time pressure. If the word is monosyllabic, the text is suboptimal for bearing a pitch movement. The insertion of a schwa in such cases enables the tune to be realised on a longer stretch of pitch-bearing material. This adjustment is further affected by the complexity and direction of the pitch movement to be realised. More complex tunes (rise-fall-rise) need more time to be realised than simple tunes (fall), thus schwa is more likely to be inserted in questions than in statements, and if it is inserted, it is longer. Likewise, rising tunes, all other things being equal, take longer to execute than falling tunes (Ohala and Ewan, 1973; Xu and Sun, 2002), thus schwa is more likely to be needed in list items bearing rising tunes (non-final and prefinal) than those bearing falling ones (final position). The pressure to insert a schwa is less acute in disyllabic words, possibly accounting for the mixed picture in the disyllabic list data set.

In addition to the effects attributable to time pressure and to properties of the tune, there was a great deal of variability in the occurrence and duration of schwa. An exploratory analysis revealed that speaker-specific patterns make the strongest contribution toward accounting for this, despite the fact that factors known to lead to speaker-specific variability were kept constant (in the current study: gender, education, regional variety spoken). One factor not directly controlled for was proficiency in English, which has been shown to play a role in vowel insertion in consonant-final loan words in Korean (Kwon, 2017). However, at the time of recording, all participants had a similar level of English (a minimum of eight years of English at school and at least one additional English course at University).

Our exploratory analyses did not reveal effects of phonetic factors, suggesting that the identity of the word final consonant does not account for the presence of schwa. This was surprising because it is well established that vowel insertions can be strongly affected by their surrounding laryngeal and supralaryngeal articulatory environment. For example, vowel insertion can be caused by misperception of word final consonant releases (e.g., Dupoux *et al.*, 1999; Kang, 2003). This misperception is known to be affected by the voicing of the consonant release, with more inserted vowels perceived after voiced consonants (e.g., Kwon, 2017, for a recent discussion on Korean). Alternatively, vowel insertion has been described as an articulatory artefact. For example, schwa in onset clusters in Tashlhiyt Berber has been found to be highly dependent on the voicing of the

consonants in the cluster (Ridouane and Fougeron, 2011). Ridouane and Fougeron (2011) conclude that schwa arises from underlap—a reduction in overlap—between the supralaryngeal constrictions for the two consonants (Steriade, 1990; Browman and Goldstein, 1992; Hall, 2006). Both articulatory and perceptual accounts imply that inserted vowels are to some extent predictable from the laryngeal specification of the consonantal environment, a diagnostic that is often associated with intrusive vowels, which, according to Hall (2006), are not considered to be phonological. Our exploratory study did not reveal any evidence for an effect of the consonant identity, suggesting that schwa insertion in Bari Italian is not affected by its segmental environment.

With these results in mind, we return to the question of how schwa in Bari Italian can be characterised in terms of Hall's (2006) typology. Is it an epenthetic vowel,<sup>5</sup> i.e., an element that has a stable form and distributions, inserted to repair illicit structures, and visible to the phonological system? The evidence we have presented suggests that some of these diagnostics match our observations. Schwa in our data is acoustically salient and surfaces frequently. It could be argued to repair illicit phonotactic structures, since consonant-final words are marginal in the native vocabulary. Is it visible to the phonological system? Our results point in this direction too, given that schwa is systematically used to realise intonational movements and is adjusted according to this functional pressure, with a greater number of schwas in monosyllabic words, especially when the tonal contour is complex or rising (rise-fall-rise, low rise, or high rise). Thus, schwa can be considered to be phonological to the extent that it is necessary for a structural description of the intonation system. This, however, does not necessarily imply that schwa is a phonological unit relevant to syllable structure (see Roettger, 2017, for a similar argumentation regarding schwa in Tashlhiyt). Our data cannot provide a conclusive answer to the question as to whether schwa is involved in building an extra syllable, making the monosyllables disyllabic, and the disyllables trisyllabic. This is even more so the case, since the addition of a further syllable would require the final consonant of the word to be geminated, as discussed in Sec. 1B. The variability mentioned in the works of Bertinetto (1985) and Repetti (2012) is confirmed in our corpus, with no clear trend toward longer consonants preceding a schwa that would be an indication of gemination.<sup>6</sup>

Despite some phonological properties discussed above, schwa in Bari Italian exhibits a large amount of variability, both within and across speakers, characteristics that are typical of intrusive vocoids, i.e., phonetic artefacts that do not have a phonological status. Comparing our findings to those on other languages with inserted schwa, we find that there are considerable differences in the factors conditioning schwa. What is striking is that, across the different studies, schwa insertion is usually affected by a combination of factors from both the linguistic and the phonetic domains. Schwa in Tashlhiyt Berber exhibits very similar patterns to Bari Italian, with schwa being determined by both tune-text-requirements and phonetic factors (Roettger, 2017). Moreover, schwa appears in Tashlhiyt to be

sociolinguistically conditioned, a factor we were unable to test in our Italian data set.

Tunisian Arabic is different from Bari Italian and Tashlhiyt Berber, in that schwa is only found in yes-no questions, that is, not at all in statements or lists. Hellmuth (2018) finds that schwa is, like in Tashlhiyt, sociolinguistically conditioned. Although schwa insertion is restricted to questions, it is only found in roughly half of the questions analysed. Hellmuth argues that it may in fact be an emerging morphological marker for interrogatives. An important difference between Bari Italian and Tunisian Arabic is that the latter language does not show any evidence of tonal crowding leading to schwa insertion, there being no observed tendency for words with final stress to insert schwa more frequently than words with stress earlier in the word. Phonetic factors did, however, play a role, with more schwa after sonorants than after obstruents (although there was still a considerable number of schwas in this environment, too).

In Standard European Portuguese (Lisbon variety), like in Tunisian Arabic, schwa is inserted in yes-no questions but not in statements. Frota (2002) points out that schwa is inserted as one of a number of accommodation strategies when the final syllable in the phrase bears a nuclear accent and ends in a sonorant. In Frota *et al.* (2016), a corpus study showed that schwa insertion (referred to as epenthesis) is found in yes-no questions not only in Lisbon, but also in the centre-southern interior regions, albeit only 17% of the time.

Thus, in both European Portuguese and Tunisian Arabic there have been reports of variation in the presence of schwa, but in both languages, unlike in Bari Italian, schwa was not found in statements or lists (although schwa was found in vocatives). In terms of possible phonetic conditioning of schwa, the properties of adjacent consonants appear to play no role in Bari Italian, unlike in the other two languages.

A further variety of European Portuguese, the Alentejo variety, is also different again, in that schwa is inserted phrase-finally after sonorants in both questions and statements. Cruz (2013) argues that schwa is inserted as a result of a following intonation phrase boundary, with no reported effect of tune. However, there is some variation conditioned by different segments (within the sonorant group) and some sociolinguistic variation, although unlike the languages and varieties discussed so far, younger speakers insert fewer schwas, interpreted as indicating that schwa insertion is in decline.

From the above brief survey, it should be clear that the insertion of schwa word finally involves variation within and across languages. In these languages, the presence of schwa is related to postlexical and metrical factors, and in Tunisian Arabic it might even be taking on a morphological status as a question affix or clitic. The status of schwa on lower prosodic levels such as syllable structure is often unclear, suggesting different degrees of phonological entrenchment. Additionally, in some of the languages and varieties discussed, there appears to be variation that could be attributed to properties of the consonant preceding the schwa, albeit to different degrees, one of the diagnostics for intrusive vowels. Although not tested for explicitly, our data provide no evidence for such an effect for schwa in Bari Italian.

All in all, not only the insertion of schwa in Bari Italian but also its insertion in other languages calls for a reframing of the typological dichotomy between intrusive and epenthetic vowels in favour of a continuum along which all of these languages can be situated. Although we have discussed cross-linguistic evidence that intonational tones play a considerable role in determining schwa insertion, it is clear that there is an interplay of different sources of this restructuring of the text.

In sum, despite effects of idiosyncratic properties of speakers and words, our findings indicate that the presence and duration of schwa in Bari Italian is driven by the functional pressure to realise communicatively relevant tonal movements (question vs statement or position in a list). In this sense, schwa cannot be considered a mere phonetic artefact, since it is relevant for phonology, in that it facilitates the production of communicatively relevant intonation contours.

## ACKNOWLEDGMENTS

This work was supported by funds for the Collaborative Research Center “1252 Prominence in Language” (German Research Council). Thank you to Mario Refice for his help with processing the speech data. This paper has benefited from very insightful comments received from Dani Byrd and an anonymous reviewer.

## APPENDIX: MODEL OUTPUT

The model output of the maximally converging models is reproduced in Table II below.

TABLE II. Model output of the maximally converging models. The tables show the estimates, SEs,  $z$ -value for logistic regressions, and  $t$ -values for linear regressions, respectively, as well as  $p$ -value based on simple Wald- $z$  tests and  $t$ -tests, respectively. These  $p$ -values differ from the  $p$ -values reported in the text. The  $p$ -values in the text are based on likelihood ratio tests. The model estimates are based on sum-to-zero contrast coded predictors and are to be interpreted as follows: The Intercept is the grand mean. In the case of the question-answer data set (a,c), the tune coefficient is the differences between the mean and questions/statements, respectively. In the case of the list data set (b,d), the tune (final) coefficient is the difference between the mean and the final condition; the tune (non-final) coefficient is the difference between the mean and the non-final condition. In turn, the coefficient of the prefinal tune condition is the difference between the mean and the sum of the final and the non-final coefficient. The metrical structure coefficient is the difference between the mean and monosyllabic/disyllabic words, respectively. The interaction coefficients indicate how much these main effects need to be adjusted across conditions.

(a) Model output for question-answer subset predicting presence of schwa				
	Estimate	SE	$z$ -value	$\Pr(> z )$
(Intercept: Mean)	3.07	0.77	4.0	0.0001
Tune	1.42	0.23	6.2	0.0000
Metrical structure	-2.07	0.68	-3.1	0.0021
Tune $\times$ metrical structure	-0.77	0.23	-3.4	0.0008

(b) Model output for list subset predicting presence of schwa				
	Estimate	SE	$z$ -value	$\Pr(> z )$
(Intercept: Mean)	1.89	0.46	4.1	<0.0001
Tune (final)	-1.18	0.21	-5.5	<0.0001

TABLE II. (Continued)

(b) Model output for list subset predicting presence of schwa				
	Estimate	SE	$z$ -value	$\Pr(> z )$
Tune (non-final)	0.81	0.35	2.3	0.0207
Metrical structure	-2.17	0.34	-6.3	<0.0001
Tune (final) $\times$ metrical structure	1.21	0.21	5.7	<0.0001
Tune (non-final) $\times$ metrical structure	-1.32	0.35	-3.7	0.0002

(c) Model output for question-answer subset predicting schwa duration				
	Estimate	SE	$t$ -value	$\Pr(> t )$
(Intercept: Mean)	90.7	5.4	16.7	<0.0001
Tune	13.4	4.1	3.3	0.0083
Metrical structure	-10.7	2.1	-5.1	0.0001
Tune $\times$ metrical structure	-6.1	1.6	-3.9	0.0027

(d) Model output for list subset predicting schwa duration				
	Estimate	SE	$t$ -value	$\Pr(> t )$
(Intercept: Mean)	81.04	3.49	23.2	<0.0001
Tune (final)	-3.25	4.42	-0.7	0.4793
Tune (non-final)	3.22	2.78	1.2	0.2613
Metrical structure	-16.77	2.63	-6.4	<0.0001
Tune (final) $\times$ metrical structure	11.03	2.12	5.2	0.0001
Tune (non-final) $\times$ metrical structure	-6.12	3.44	-1.8	0.0953

<sup>1</sup>Many tone languages restrict contour tones to syllables with rhymes that contain more sonorous elements (Zhang, 2004; Gordon, 2004), suggesting a relation between tonal configuration and the segments that bear them regardless of the source of the tone.

<sup>2</sup>We acknowledge that the design exhibits an imbalance between monosyllables and disyllables. This imbalance was an artefact of the corpus being borrowed from an earlier study on Bari Italian. While an asymmetric number of items across conditions is not ideal, the statistical models fitting the data are not directly affected by this asymmetry.

<sup>3</sup><https://osf.io/2n6bj/> (Last viewed April 7, 2018).

<sup>4</sup>To increase readability, we will only report respective  $p$ -values in the text. Estimates and margins of errors are given in Figs. 4 and 5, descriptive means are given in Table I. The model output is reproduced in the Appendix. The data table and all R scripts are available online: <https://osf.io/2n6bj/> (Last viewed April 7, 2018).

<sup>5</sup>Word final epenthetic vowels are more accurately referred to as paragogic vowels. We adhere to Hall’s terminology in the current study.

<sup>6</sup>Due to the unbalanced distribution of schwa across speakers, target words, and prosodic conditions, neither reliable inferential nor descriptive assessments of consonantal duration can be performed in our data set.

<sup>7</sup>We transcribe the intervocalic nasal in *Dennis* as long to indicate its geminate status. We do not, however, transcribe length on the word final consonants preceding a schwa. See Sec. IV for a discussion of the status of these consonants.

Abercrombie, D. (1967). *Elements of General Phonetics* (Edinburgh University Press, Edinburgh).

Bannert, R., and Bredvad-Jensen, A. (1975). “Temporal organisation of Swedish tonal accent: The effect of vowel duration,” *Work. Pap. Linguist., Lund Univ.* **10**, 1–36.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). “Fitting linear mixed-effects models using lme4,” *J. Stat. Software* **67**, 1–48.

Bertinetto, P. M. (1985). “A proposito di alcuni recenti contributi alla prosodia dell’italiano” (“About some recent contributions to the prosody of Italian”), *Annali della Scuola Normale Superiore di Pisa. Classe di Lettere e Filosofia* **15**, 581–643.

Bolinger, D. L. (1957). *Interrogative Structures of American English (The Direct Question)*, publication of the American Dialect Society (University of Alabama Press, Tuscaloosa, AL), p. 28.

- Boersma, P. (2001). "Praat, a system for doing phonetics by computer," *Glott International* 5(9/10), 341–345.
- Breiman, L. (2001). "Random forests," *Mach. Learn.* 45, 5–32.
- Broniś, O. (2016). "Italian vowel paragoge in loanword adaptation. Phonological analysis of the Roman variety of Standard Italian," *Ital. J. Linguist.* 28(2), 25–68.
- Browman, C. P., and Goldstein, L. (1992). "'Targetless' schwa: An articulatory analysis," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. J. Docherty and D. R. Ladd (Cambridge University Press, Cambridge), pp. 26–56.
- Bruce, G. (1977). "Swedish word accents in sentence perspective," Ph.D. dissertation, Lund University, Sweden.
- Cangemi, F., and Grice, M. (2016). "The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation," *Lab. Phonol.* 7, 1–9.
- Caspers, J., and van Heuven, V. J. (1993). "Effects of time pressure on the phonetic realization of Dutch accent-lending pitch rise and fall," *Phonetica* 50, 161–171.
- Cruz, M. (2013). "Prosodic variation in European Portuguese: Phrasing, intonation and rhythm in central-southern varieties," Ph.D. dissertation, Universidade de Lisboa, Portugal.
- D'Achille, P. (2010). *L'Italiano Contemporaneo* (Contemporary Italian), Il Mulino, Bologna, Italy.
- Dell, F., and Elmedlaoui, M. (2013). "Syllables and gemination in imperfective stems in Tashlhiyt Berber," *Brill's J. Afroasiatic Lang. Linguist.* 5, 1–34.
- D'Imperio, M., and House, D. (1997). "Perception of questions and statements in Neapolitan Italian," in *Proceedings of Eurospeech '97*, Vol. 1, pp. 251–254.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). "Epenthetic vowels in Japanese: A perceptual illusion?," *J. Exp. Psychol.: Human Percept. Perform.* 25, 1568–1578.
- Erikson, Y., and Alstermark, M. (1972). "Fundamental frequency correlates of the grave word accent in Swedish: The effect of vowel duration," *Speech Transm. Lab. Quart. Papers Status Rep.* 13(2–3), 53–60.
- Frota, S. (2002). "Tonal association and target alignment in European Portuguese nuclear falls," in *Papers in Laboratory Phonology 7*, edited by N. Warner and C. Gussenhoven (Mouton de Gruyter, Berlin, New York), Vol. 7, pp. 387–418.
- Frota, S. (2014). "The intonational phonology of European Portuguese," in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, Oxford), pp. 6–42.
- Frota, S., Cruz, M., Castelo, J., Barros, N., Crespo-Sendra, V., and Vigário, M. (2016). "Tune or text? Tune-text accommodation strategies in Portuguese," in *Proceedings of the Speech Prosody 2016*.
- Gartenberg, R., and Panzlaff-Reuter, C. (1991). "Production and perception of F0 peak patterns in German," *Arbeitsber. Inst. Phonetik Univ. Kiel* 25, 29–115.
- Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M., and Sorianello, P. (2015). "Varieties of Italian and their intonational phonology," in *Intonation in Romance*, edited by S. Frota and P. Prieto (Oxford University Press, Oxford), pp. 140–197.
- Gilles, P. (2005). *Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung* (Regional Prosody of German: Variability in the Intonation of Terminality and Continuation) (Walter de Gruyter, Berlin).
- Goldsmith, J. A. (1976). *Autosegmental Phonology* (Indiana University Linguistics Club, Bloomington), Vol. 159.
- Gordon, M. (2004). "Syllable weight," in *Phonetic bases for phonological markedness*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, Cambridge), pp. 277–312.
- Grabe, E. (1998). "Pitch accent realization in English and German," *J. Phonetics* 26, 129–143.
- Grabe, E., Post, B., Nolan, F., and Farrar, K. (2000). "Pitch accent realization in four varieties of British English," *J. Phonetics* 28, 161–185.
- Grice, M. (1995). *The Intonation of Interrogation in Palermo Italian: Implications for Intonation Theory* (Niemeyer, Tübingen, Germany).
- Grice, M., Ridouane, R., and Roettger, T. B. (2015). "Tonal association in Tashlhiyt Berber: Evidence from polar questions and contrastive statements," *Phonology* 32, 241–266.
- Grice, M., Savino, M., and Refice, M. (1997). "The intonation of questions in Bari Italian: Do speakers replicate their spontaneous speech when reading," *Phonon* 3, 1–7.
- Grønnum, N. (1989). "Stress group patterns, sentence accents and sentence intonation in Southern Jutland (Sonderborg and Tønder)—With a view to German," *Annu. Rep. Inst. Phonetics, Univ. Copenhagen* 23, 1–85.
- Hall, N. (2006). "Cross-linguistic patterns of vowel intrusion," *Phonology* 23, 387–429.
- Hanssen, J. (2018). "Regional variation in the realization of intonation contours in the Netherlands," Ph.D. dissertation, Utrecht University, Utrecht.
- Harms, R. T. (1976). "The segmentalization of Finnish 'nonrules,'" *Texas Linguist. Forum* 5, 73–88.
- Hellmuth, S. (2018). "Text-tune alignment in Tunisian Arabic yes-no questions," in *Prosodic Variation (With)in Languages: Intonation, Phrasing and Segments*, edited by M. Cruz, S. Frota, and P. Oliveira (Equinox, London) (in press).
- Heston, T. M. (2014). "Prosodic differences between declaratives and polar questions in Fataluku," in *28th Pacific Asia Conference on Language, Information and Computation*, pp. 395–403.
- Kang, Y. (2003). "Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean," *Phonology* 20, 219–273.
- Kilbourn-Ceron, O., and Sonderegger, M. (2018). "Boundary phenomena and variability in Japanese high vowel devoicing," *Nat. Lang. Linguist. Theory* 36(1), 175–217.
- Klajn, I. (1972). *Influssi inglesi nella lingua italiana (English Influences on the Italian Language)* (Olschki, Firenze).
- Krämer, M. (2009). *The Phonology of Italian* (Oxford University Press, Oxford).
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). "lmerTest package. Tests in linear mixed effects models," *J. Stat. Software* 82(13), 1–26.
- Kwon, H. (2017). "Language experience, speech perception and loanword adaptation: Variable adaptation of English word-final plosives into Korean," *J. Phonetics* 60, 1–19.
- Ladd, D. R. (2008). *Intonational Phonology*, 2nd ed. (Cambridge University Press, Cambridge).
- Leben, W. R. (1976). "The tones in English intonation," *Linguist. Anal.* 2, 69–107.
- Lepschy, A. L., and Lepschy, G. (1992). *The Italian Language Today* (Routledge, London).
- Levin, J. (1987). "Between epenthetic and excrescent vowels," in *Proceedings of the 6th West Coast Conference on Formal Linguistics*, University of Arizona, Vol. 6, pp. 187–202.
- Liberman, M. Y. (1975). "The intonational system of English" Doctoral dissertation, MIT.
- Lickley, R. J., Schepman, A., and Ladd, D. R. (2005). "Alignment of 'phrase accent' lows in Dutch falling rising questions: Theoretical and methodological implications," *Lang. Speech* 48, 157–183.
- Lüdecke, D. (2017). *ggeffects: Create Tidy Data Frames of Marginal Effects for 'ggplot' from Model Outputs*, R package version 0.1.2, <https://CRAN.R-project.org/package=ggeffects>.
- Mücke, D., Grice, M., Becker, J., and Hermes, A. (2009). "Sources of variation in tonal alignment: Evidence from acoustic and kinematic data," *J. Phonetics* 37, 321–338.
- Ohala, J. J., and Ewan, W. G. (1973). "Speed of pitch change," *J. Acoust. Soc. Am.* 53, 345.
- Peng, R. D. (2011). "Reproducible research in computational science," *Science* 334, 1226–1227.
- Peters, J. (2006). "Intonation deutscher Regionalsprachen" ("Intonation of German regional varieties"), in *Linguistische Impulse und Tendenzen (Linguistic Impulses and Tendencies)* (Walter de Gruyter, Berlin).
- Pierrehumbert, J. (1980). "The phonology and phonetics of English intonation," Doctoral dissertation, MIT, Cambridge, MA.
- Pierrehumbert, J., and Beckman, M. (1988). *Japanese Tone Structure* (MIT Press, Cambridge, MA).
- Prieto, P., D'Imperio, M., and Gili Fivela, B. (2005). "Pitch accent alignment in romance: Primary and secondary associations with metrical structure," *Lang. Speech* 48, 359–396.
- Prieto, P., and Ortega-Llebaria, M. (2009). "Do complex pitch gestures induce syllable lengthening in Catalan and Spanish," in *Phonetics and Phonology: Interactions and Interrelations*, edited by M. Vigário, S. Frota, and M. J. Freitas (Benjamins, Amsterdam), pp. 51–70.
- Prieto, P., van Santen, J., and Hirschberg, J. (1995). "Tonal alignment patterns in Spanish," *J. Phonetics* 23, 429–451.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

- Rathcke, T. (2009). *Komparative Phonetik und Phonologie der Intonationssysteme des Deutschen und Russischen* (Comparative phonetics and phonology of the intonation systems of German and Russian) (Herbert Utz Verlag, München), Vol. 29.
- Refice, M., Savino, M., and Grice, M. (1997). "A contribution to the estimation of naturalness in the intonation of Italian spontaneous speech," in *Proceedings of the 5th European Conference on Speech Communication and Technology*, Rhodes, Greece, pp. 783–786.
- Repetti, L. (1993). "The integration of foreign loans in the phonology of Italian," *Italica* 70(2), 182–196.
- Repetti, L. (2012). "Consonant-final loanwords and epenthetic vowels in Italian," *Catalan J. Linguist.* 11, 167–188.
- Ridouane, R., and Fougeron, C. (2011). "Schwa elements in Tashlhiyt word-initial clusters," *J. Lab. Phonol.* 2, 275–300.
- Roettger, T. B. (2017). *Tonal Placement in Tashlhiyt: How an Intonation System Accommodates to Adverse Phonological Environments* (Language Science, Berlin).
- Roseano, P., del Mar Vanrell, M., and Prieto, P. (2015). "Intonational phonology of Friulian and its dialects," in *Intonational Variation in Romance*, edited by S. Frota and P. Prieto (Oxford University Press, Oxford), pp. 101–139.
- Savino, M. (2001). "Non-finality and pre-finality in Bari Italian intonation: A preliminary account," in *Proceedings of the VII European Conference on Speech Communication and Technology*, Aalborg, pp. 939–942.
- Savino, M. (2004). "Intonational cues to discourse structure in a variety of Italian," in *Regional Variation in Intonation*, edited by P. Gilles and J. Peters (Niemeyer, Tuebingen), pp. 145–159.
- Savino, M. (2012). "The intonation of polar questions in Italian: Where is the rise?," *J. Int. Phonetic Assoc.* 42, 23–48.
- Savino, M., Grice, M., Gili Fivela, B., and Marotta, G. (2006). "Intonational cues to discourse structure in Bari and Pisa Italian: Perceptual evidence," in *Proceedings of SPEECH PROSODY 2006*, Dresden, pp. 114–117.
- Schepman, A., Lickley, R., and Ladd, D. R. (2006). "Effects of vowel length and 'right context' on the alignment of Dutch nuclear accents," *J. Phonetics* 34, 1–28.
- Silverman, D. (2011). "Schwa," in *The Blackwell Companion to Phonology*, edited by M. van Oostendorp, C. J. Ewen, E. V. Hume, and K. Rice (Wiley-Blackwell, Malden, MA), pp. 628–642.
- Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2017). *afex: Analysis of Factorial Experiment*, R package version 0.16-1, <https://CRAN.R-project.org/package=afex>.
- Steele, S. A. (1986). "Nuclear accent F0 peak location: Effects of rate, vowel, and number of following syllables," *J. Acoust. Soc. Am.* 80, S51.
- Steriade, D. (1990). "Gestures and autosegments: Comments on Browman and Goldstein's paper," in *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston, and M. E. Beckman (Cambridge University Press, Cambridge), pp. 382–397.
- Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., and Zeileis, A. (2008). "Conditional variable importance for random forests," *BMC Bioinformatics* 9, 307.
- Tagliamonte, S. A., and Baayen, R. H. (2012). "Models, forests, and trees of York English: Was/were variation as a case study for statistical practice," *Lang. Var. Change* 24, 135–178.
- Warner, N., Jongman, A., Cutler, A., and Mücke, D. (2001). "The phonological status of Dutch epenthetic schwa," *Phonology* 18, 387–420.
- Wickham, H. (2017). "Tidyverse: Easily install and load 'Tidyverse' packages," R package version 1.1.1. <https://CRAN.R-project.org/package=tidyverse>.
- Winter, B., and Grawunder, S. (2012). "The phonetic profile of Korean formal and informal speech registers," *J. Phonetics* 40, 808–815.
- Xu, Y., and Sun, X. (2002). "Maximum speed of pitch change and how it may relate to speech," *J. Acoust. Soc. Am.* 111, 1399–1413.
- Zhang, J. (2004). "The role of contrast-specific and language-specific phonetics in contour tone distribution," in *Phonetically Based Phonology*, edited by B. Hayes, R. M. Kirchner, and D. Steriade (Cambridge University Press, Cambridge), pp. 157–190.