



BRILL



brill.com/ldc

The tune drives the text

Competing information channels of speech shape phonological systems

Timo B. Roettger

Northwestern University

timo.b.roettger@gmail.com

Martine Grice

University of Cologne

martine.grice@uni-koeln.de

Abstract

In recent years there has been increasing recognition of the vital role of intonation in speech communication. While contemporary models represent intonation—the tune—and the text that bears it on separate autonomous tiers, this paper distils previously unconnected findings across diverse languages that point to important interactions between these two tiers. These interactions often involve vowels, which, given their rich harmonic structure, lend themselves particularly well to the transmission of pitch. Existing vowels can be lengthened, new vowels can be inserted and loss of their voicing can be blocked. The negotiation between tune and text ensures that pragmatic information is accurately transmitted and possibly plays a role in the typology of phonological systems.

Keywords

intonation – epenthesis – devoicing – lengthening – phonological typology

1 Introduction

During the act of speech, speakers simultaneously express multiple dimensions of meaning. For example, a simple sentence such as “The tune drives the text” expresses the proposition at hand through its lexical concepts and their morphosyntactic combination. Beyond that, other channels of the speech signal,

such as intonation, can express pragmatic functions such as illocutionary force and information structure.

Traditionally, much attention has been paid to lexical meaning, as expressed through consonants and vowels organized into metrical structures, referred to as the *text*. More recently, research has been increasingly focused on utterance-level meanings as expressed by intonation, referred to as the *tune*. Utterance-level meaning expressed by intonation involves paralinguistic functions as diverse as emotion, attitude and speaker involvement, and pragmatic functions, such as the illocutionary force behind a proposition like declaring the proposition or asking for confirmation (e.g. “The tune drives the text.” vs. “The tune drives the text?”) or information structure, i.e. contextualizing the proposition within the immediate discourse (e.g. “The TUNE drives the text.” vs. “The tune drives the TEXT.”, as answers to “What drives the text?” and “What does the tune drive?” respectively).

The intonation contour of an utterance is traditionally thought to be independent of the words that bear it (e.g. Abercrombie, 1967; Pierrehumbert, 1980; Ladd 2008), the two being represented on separate levels, allowing them to be freely combinable: “not only can the same text have many different melodies, the same melody can occur on many different texts” (Pierrehumbert, 1980:8). The separation of tune and text has led to important insights into how prosodic systems are structured, how intonational contours are aligned with the available segmental structure, and how intonational meaning is expressed. However, accumulating empirical evidence has made it clear that segmental and intonational aspects of speech are inherently connected. For intonation to be produced, voiced segmental material is needed to enable the vocal folds to vibrate; for intonation to be perceived, it is important for the segmental material to exhibit periodic energy with a rich harmonic structure (Zhang, 2004; Barnes et al., 2014). Thus, amongst the segments that make up the text, some do not allow for vocal fold modulation at all, and others are not sufficiently periodic for optimal perceptual retrieval of such modulation. The shortage of pitch bearing segmental material can lead to cases in which a communicatively relevant tonal contour (e.g. a pitch movement distinguishing a question from a statement) might be phonetically impoverished, potentially obscuring a pragmatic contrast that would otherwise be signaled by intonation. As we will see in the body of this paper, such cases are cross-linguistically very common.

The present paper will show that languages resolve this functional conflict in different ways. In section 2, we briefly introduce traditional models of intonational phonology that propose separate levels of representation for the text and the tune. We will then summarize cross-linguistic evidence suggesting that functional conflicts between signaling lexical and intonational meaning

are common, and when they happen, linguistic systems have been shown to accommodate such functional conflicts in different ways. Some of these lead to adjustments of the tune (the text drives the tune, section 3), others lead to the adjustment of the text (the tune drives the text, section 4). The latter case has important implications for phonological typology and should help us further our understanding of why sound systems are the way they are (section 5).

2 Models of intonation

Earlier work on intonation distinguished between tune and text by representing the tune as diacritics on and around the orthographic or phonetic transcription of the consonants and vowels, representing the text (Trager & Smith, 1951; Halliday, 1967; Crystal, 1969; Tench, 1996). Although these early models still considered the tune—in the form of phonological tones—to be a property of segments, it was acknowledged that tones could stretch over multiple segments, leading to the term *suprasegmental*. However, work on a number of African languages in which tones have lexical and grammatical functions revealed that tones could not only span more than one segment but that the very same segments could bear different tones in different lexical and grammatical contexts (Leben, 1973). Moreover, tones expressing certain grammatical functions might not have a host segment at all, leading to a situation where the tone is floating. Floating tones can also result from deletion of a segment, either diachronically or synchronically. This apparent autonomy of the tune called for a new model which could account for the relation between tune and text.

The new model was *Autosegmental Phonology* (Goldsmith, 1976). In this model, tones are not properties of segments. Instead, they are treated as autonomous units on a par with segments, but represented on their own tier. Not every segment is a tone-bearing unit (TBU); initially this function was given to vowels, and later to the syllable (Kahn, 1976) of which the vowel was the nucleus. The link between the segmental tier and the tonal tier is represented as association lines linking tones to TBUs. These association lines can be deleted, just as new lines can be added, allowing tones to spread to TBUs that have no tone, operations that are subject to universal well-formedness conditions. This architecture allows the tune to be represented independently of the text, yet provides a mechanism to capture the phenomena observed in a reasonably constrained way.

The autosegmental model was applied to post-lexical tones, i.e. intonation (Goldsmith, 1974; Liberman, 1975; Leben, 1976), in particular to English, in which tones are more sparsely distributed than in the languages previously

analyzed with this model. Goldsmith (1974) showed that the same tune could be produced on strings of different lengths. Thus, the tonal sequence H*L representing a falling intonation typical of neutral statements could be produced on words of varying length and structure, such as *ANN*, *ANabel* and *JoANNE*, as in (1). However, it is not sufficient to indicate that the fall occurs on a particular word; its location within the word is also important. This is indicated by a 'star', marking a metrically strong syllable for association with the tonal sequence. Moreover, the timing of the fall in relation to the metrically strong (starred) syllable could be captured by a star on one of the two tones. In (1) the star is on the H tone, indicating that the fall begins on, rather than before, the starred syllable.

- (1) H*L H*L H*L
 | | |
 * * *
- ANN ANabel JoANNE

Thus, in autosegmental models of English, the tune and the text were on independent tiers but a mechanism was put in place not only to associate tones with words but also to privilege certain tones and certain TBUS (using the star notation) to ensure an appropriate association, even if the number of tones and TBUS was not equal on each level.

A further development involved the structuring of the text into higher level constituents, such that TBUS were designated terminal elements of metrical trees. *Autosegmental-metrical Phonology* (Pierrehumbert, 1980; Ladd, 2008) developed from these two approaches, providing a framework for representing associations to TBUS, represented as the heads of certain constituents, and to edges of those and higher constituents. However, these two tiers naturally interact with each other.

For a phonological tone to be realized as a perceived pitch contour, it is generally agreed that it needs to be realized on segmental material that is voiced. For the contour to be perceived effectively, it needs to occur on segmental material of high periodic energy (Zhang, 2004; Barnes et al., 2014). However, some of the segments used to express lexical meaning are well suited neither to producing pitch modulation nor to enabling a perceptual retrieval of such modulation. The acoustic correlate of perceived pitch is fundamental frequency (f_0), which can only be reliably calculated if there is sufficient strength in the periodic energy of the speech signal.

Moreover, there is a tendency for multiple tones to occur at edges of constituents (e.g. if the head is close to an edge), a scenario that can lead to

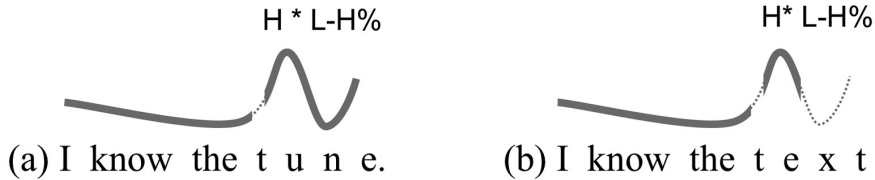


FIGURE 1 Schematic representation of the f_0 contour of the English utterances (a) “I know the tune.” and (b) “I know the text.”, indicating the meaning of uncertainty, expressed by a rise-fall-rise ($H^* L-H\%$) on both phrases. In (b), significant parts of the final rise-fall-rise would be required to co-occur with voiceless segments. Voiceless portions of the signal indicated by the thin dotted line.

EXAMPLE CONSTRUCTED BY THE 2ND AUTHOR

so-called tonal crowding, in which a number of different tones are all simultaneously associated with one syllable. This multiple association leads to a sequencing of the tones (e.g. LHLH) representing a dynamic intonation pattern (in this case rise-fall-rise). Depending on the makeup of the syllable, it may not be possible for the entire movement to be realized—all other things being equal. In such cases, important acoustic information for the recognition of the speaker intention could be lost. Consider the phrases “I know the tune.” and “I know the text.”, both with an intonation pattern indicating uncertainty (Ward & Hirschberg, 1985), hypothetical pitch contours of which are illustrated in Fig. 1.

Whereas the rhyme of the final word in (a), ‘tune’, is fully voiced, the rhyme of the final word in (b), ‘text’, is composed of a short vowel and a voiceless coda cluster. This segmental make-up heavily constrains the realization and perceptual retrieval of intonational movements, especially when they are complex, such as in the rise-fall-rise. This scenario marks a functional conflict, potentially endangering a pragmatic contrast signaled by intonation. In the following two sections, we will see how such conflicts can be resolved.

3 The text drives the tune

One way to resolve a tune-text conflict is to adjust the tune (Erikson & Alstermark, 1972; Bannert & Bredvad-Jensen, 1975; Grønnum, 1989; Grice, 1995; Grabe, 1998; Grabe et al., 2000; Lickley, Schepman, & Ladd, 2005; Rathcke, 2009; Grice et al., 2015). The nature of these adjustments depends on syntagmatic, paradigmatic, and language-specific factors (see Hanssen, 2017; or Roettger, 2017, for recent discussions on how the text can drive the tune).

Adjustments to the tune were first reported for Swedish, a language that has two lexically distinctive accent categories. Both accent types, Accent 1 and

Accent 2, are characterized by a falling pitch contour, represented as a sequence of high and low tones, differing only in the alignment of the high tone: the high tone in Accent 1 words reaches its target earlier than the high tone in Accent 2 words (Bruce, 1977). Erikson and Alstermark (1972) discussed the adjustment of these accent patterns in light of the segmental structure of the words bearing them. On the one hand, they show that f_0 movement is reduced with decreasing vowel length. In this case, the tonal target is undershot, with the fall not reaching what is assumed to be its low target compared to words with sufficient sonorous segments. In the most extreme case, e.g. if the syllable contains a short vowel and a voiceless coda consonant, there may not remain any trace of a fall in f_0 . This adjustment to the tune is called “truncation”. Alternatively, the full pitch movement is sometimes realized more rapidly on shorter vowels. These local “rate adjustments” were subsequently called “compression” by Bannert and Bredvad-Jensen (1975).

These two adjustments to the tune, truncation and compression, have since been attested for intonational tune-text conflicts across many languages. Grønnum (1989) reports truncation for Danish rise-falls and Northern German falls. Grabe (1998) reports on truncation and compression for standard varieties of German and English: In her study, Grabe systematically manipulated the segmental material on which speakers had to produce different intonational contours, ranging from a disyllabic word with a long vowel (e.g. Sheaffer /ʃi:fə/) through a monosyllabic word with a long vowel (e.g. sheaf /ʃi:f/) to a monosyllabic word with a short vowel (e.g. shift /ʃift/). Because the consonants in her stimuli were voiceless, the phonetic opportunity for realizing tonal movements was restricted to the vowel(s). Her results indicate that Southern Standard British English compresses both falls and rises, whereas Northern Standard German truncates falls and compresses rises. This is in line with truncation reports for high fall-rise contours in German and Dutch (Lickley, Schepman, & Ladd, 2005; Ladd, 2008) in which the falling component of the contour may be missing, resulting in a simple high rise. A series of studies by Rathcke (2009) demonstrates truncation for rising-falling contours in German and partial compression of falling contours. Although often reported on as two separate strategies, truncation and compression have been found to involve mixed strategies, with tunes being to some extent both truncated and compressed (Prieto & Ortega-Llebaria, 2009; Roettger, 2017).

In some cases, truncated patterns have been described as having a different representation from non-truncated patterns. For example, Grice (1995) discusses truncation in Palermo Italian in cases where there is tonal crowding: Complex rising-falling contours entirely lack their falling component if the final syllable carries a pitch accent (see Fig. 2).



FIGURE 2 Schematic representation of the f_0 contours of the Palermo Italian utterances (a) “Glielo porta domani” ‘Are you bringing it to her tomorrow?’ and (b) “Gliel’hai detto tu?” ‘Did you tell it to him?’. In (a) where stress is penultimate in the final word “domani”, the rise fall is realized, whereas in (b) where stress is on the final syllable of the phrase, there is only a rise. Voiceless portions of the signal indicated by the thin dotted line.

EXAMPLE CONSTRUCTED FROM F_0 CONTOURS IN GRICE (1995)

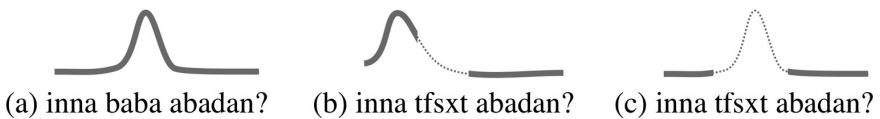


FIGURE 3 Schematic representation of the f_0 contours of the Tashlhiyt Berber utterances (a) “inna baba abadan” ‘he always said ‘father.’, and (b) and (c) “inna tfsxt abadan” ‘he always said ‘you canceled’’, expressing contrastive focus on baba and tfsxt, respectively. In (b) and (c) the contrasted word contains no voiced segments. In (b) the tonal complex is anticipated and thus realized on the preceding word, in (c) the tonal complex is lost. Voiceless portions of the signal indicated by the thin dotted line.

EXAMPLES ADAPTED FROM ROETTGER (2017)

Grice, Ridouane, and Roettger (2015) discuss tune-text conflicts in Tashlhiyt Berber, a language that allows exceptionally long sequences of voiceless consonants. One of the observed conflict resolutions is a complete truncation of a rise-fall, resulting in a loss of the entire tonal complex (see Fig. 3c).

Alternative strategies to deal with conflicts in tune-text-association involve temporal shifts of tonal targets. If the segmental material is insufficient to realize a sequence of tonal targets, they may occur instead on more suitable text (Grice et al., 2015; Roettger, 2017). Thus, instead of losing the entire tonal complex, Tashlhiyt speakers, for example, have the option of placing the tones on a preceding word that contains sufficiently sonorous segmental material for their realization (see Fig. 3b). The tendency of tones to be anticipated in cases of tonal crowding at phrase edges has been observed in a variety of languages (Steele, 1986; Caspers & van Heuven, 1993; Prieto, van Santen, & Hirschberg, 1995; D’Imperio, 2001; Prieto, D’Imperio, & Fivela, 2005; Schepman, Lickley, & Ladd, 2006; Hanssen et al., 2007; Mücke et al., 2009). For example, in Neapolitan Italian, questions and statements can be expressed with similar

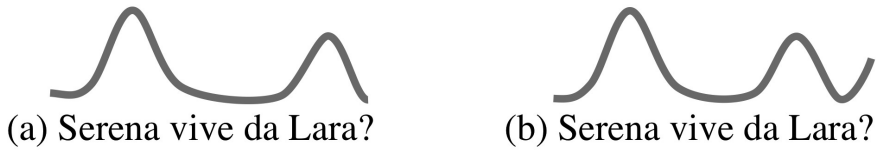


FIGURE 4 Schematic representation of the f_0 contour of the Neapolitan Italian utterance “Serena vive da Lara” ‘Serena lives at Lara’s’, expressing object focus on Lara in an interrogative with (a) a simple rise-fall on Lara, and (b) a rise-fall-rise starting before Lara.

EXAMPLES ADAPTED FROM CANGEMI & GRICE (2016: FIGURE 2)

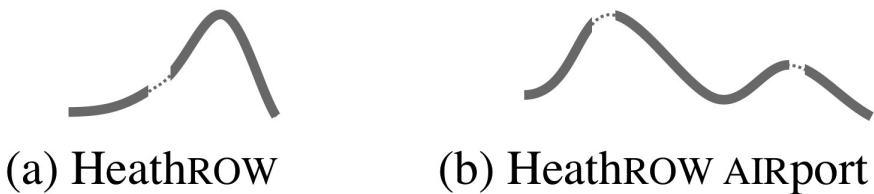


FIGURE 5 Schematic representation of the f_0 contour of the English noun phrases (a) “Heathrow” and (b) “Heathrow airport”, in which Heathrow co-occurs with a pitch accent on the primary stressed syllable (ROW) in (a) and with a pitch accent on the secondary stressed syllable (Heath) in (b). Voiceless portions of the signal indicated by the thin dotted line.

EXAMPLES CONSTRUCTED BY THE 2ND AUTHOR

tonal events, i.e. a rise-fall in pitch phrase finally (if the final word is in narrow focus). These tonal events usually differ in the alignment of the high tone, with the high tone being reached later in questions than in statements (D’Imperio & House, 1997). However, questions can also be produced with an additional final rise, leading to a more complex tonal sequence (a rise-fall-rise, as in Fig. 4b). Looking at the alignment of the high target of the pitch accent across instances with and without this utterance-final rise, the high target shifts to the left when a phrase-final rise is present (Cangemi & Grice, 2016, see Fig. 4).

Beyond cases of tonal crowding at phrase edges, tones can also be shifted in so-called stress clash contexts. For example, in words such as “Heathrow”, the part of London in which the largest airport is located, the primary stress is located on the final syllable, which usually functions as the TBU for a pitch accent. When this word is directly followed by a word with primary stress on the initial syllable (e.g. “Airport”), however, the first syllable is promoted to TBU status, so as to bear an accent (cf. Fig. 5, e.g. Bolinger, 1965; Shattuck-Hufnagel, Ostendorf, & Ross, 1994).

Such shifts in accent placement have been observed across unrelated languages (see Tilsen, 2012, for an overview). All of the interactions between tune

and text that we have discussed so far, i.e. truncation, compression, and tonal shift, can be considered negotiations between competing meaning-signaling channels of speech. In the case of truncation and compression, the information channel that is responsible for intonational meaning, the f_0 contour, is impoverished to a certain extent. Truncating or compressing the f_0 contour leads to loss of information and results in weaker perceptual retrievability of the pragmatic contrast (Odé, 2005), thus reducing the probability of the communicative functions being transmitted accurately. Tonal shift, on the other hand can be seen as upholding the intonational channel, with the caveat that a shifted tonal event may be misinterpreted as signaling meaning that was not intended (e.g., in the case of Tashlhiyt, shifting a focus-marking pitch movement to a constituent that is not focused; or giving the listener the impression that the primary stress is on a different syllable in the word, possibly impeding lexical access, and at worst activating a different lexical item from the one intended).

4 The tune drives the text

The idea that the text affects the tune is not new. There is a substantial literature on consonant-tone interactions in lexical tone systems (e.g. Bradshaw, 1999; Yip, 2002; Tang, 2008, for surveys). For example, there are strong statistical relationships between tone height and laryngeal properties of prevocalic consonants, which has been the driving force behind diachronic models of tonogenesis (e.g. Haudricourt, 1954; Hombert et al., 1979). The converse pattern, i.e. the impact of tone on its segmental environment has been largely ignored. In fact, it has been claimed that lexical tone does not affect consonants at all (Hyman, 1973, 1976). Other authors have argued that tone can have a systematic influence on consonants (Maddieson, 1974, 1976; Gandour & Maddieson, 1976). However, adjusting the text to create a segmental environment that is better suitable for realizing f_0 movements is an effective way to decrease the functional impact of tune-text conflicts. In the remainder of this paper, we will discuss evidence indicating that the realization of post-lexical tonal movements can be systematically correlated with the adjustment of the segments involved.

Although such adjustments of the text have received less attention in the literature, there are a number of languages exhibiting such patterns in phrase-final position. Typically, these adjustments are observed when there is a complex intonation contour, involving at least one rising element, and when this contour is either on a monosyllabic word or on a word with final stress, leading to tonal crowding. In the following we examine three cases in which segmen-

tal patterns have been reported to adjust. In all cases, this adjustment involves vowels, elements of high intensity and rich harmonic structure, making them, from a perceptual standpoint, the ideal carriers of pitch information: (i) Existing vowels can be lengthened, (ii) new vowels can be inserted or (iii) the devoicing or deletion of existing vowels can be blocked. It is argued that all three modifications make the segmental string more suitable for the production of pitch movements. They also make the pitch contour easier to perceive, facilitating a more robust retrieval of intonational meaning.

4.1 *Vowel lengthening*

One way to create a segmental environment that is better suited for realizing tonal movements is lengthening of existing segments that are already good carriers of the tune. Such lengthening has been observed for Bari Italian polar questions, which are typically realized with an accentual rise followed by a fall-rise. Phrase-final accented syllables bearing the rise-fall-rise are considerably lengthened in comparison to the same syllable in neutral statements, characterized by a simple fall (Grice, Savino, & Refice, 1997; Refice, Savino, & Grice, 1997). Frota (2002, 2014) reports on similar lengthening of phrase-final vowels in fall-rises in European Portuguese.¹ Lengthening has also been observed for different varieties of German (Gartenberg & Panzlaff-Reuter, 1991; Gilles, 2005). For example, Gilles (2005) found that monosyllables are longer with a fall-rise intonation than with falling contours. Ladd (2008: 183) discusses rise-fall tunes in Hungarian questions, which can only be realized in monosyllabic words when they are accompanied by prolongation of the segmental material. Similar observations have been made for Spanish and Catalan by Prieto and Ortega-Llebaria (2009). They found that, in words with final stress, syllables are longer in rise-falls than in falls. Biondi (2011) reports on unusual phrase-final vowel lengthening in accusatory speech, characterized by high pitch in Kashibo-Kakataibo, a Panoan language spoken in Peru. Vanrell and Cabré (2011) and Cabré and Vanrell (2013) report that the vocative intonation in Sardinian, which is characterized by a rise-fall pattern, is accompanied by lengthening in monosyllabic words. Heston (2014) reports on lengthening of utterance-final vowels when they are accompanied by a final rise-fall in Fataluku, a Papuan language of East Timor. Heston notes that in these cases, vowels are approximately twice as long as comparable vowels accompanied by a simple fall.

1 Vigário (2016) also raises the possibility that pitch accent assignment might trigger diphthongization in Portuguese, another possible instance of tune-driven segmental adjustment.

In all these cases, there is lengthening when the phrase-final syllable is associated with high pitch and all of the above cases are concerned with lengthening at or near prosodic edges. Similarly, lengthening has also been observed for scenarios in which pitch accents are close together, resulting in a sequence of tones in close proximity. In such contexts, the segmental material can be lengthened, resulting in more space between consecutive tonal targets (Lieberman, 1975; Tilsen, 2012).

These durational adjustments are often gradient in nature and consequently considered phonetic. However, it is also possible to interpret these adjustments as the insertion of metrical beats (Selkirk, 1984). For example, in Sanskrit, vowels are sometimes orthographically marked as ‘overlong’ (referred to as ‘pluti’, marked with the numeral ३), indicating that they consist of three moras. These overlong vowels have been described as occurring mainly at the right edge of polar questions and are associated with a particular pitch profile (Strunk, 1983). This example not only suggests a relationship between questions and vowel length but also a categorical lengthening pattern that has necessitated orthographic marking.

All of these cases, observed for a wide range of languages, can be interpreted as adjustments of the text to enable the realization of communicatively relevant pitch movements.

4.2 *Vowel insertion*

Another way to create a segmental environment that is more suitable for realizing tonal movements is by inserting a vowel. In the following, we use the term *vowel insertion* to loosely refer to any addition of non-lexical vowels including both ‘intrusive vocoids’ and ‘epenthetic vowels’, two proposed typological categories that mainly differ with respect to their status within the phonological system of the respective language (e.g. Harms, 1976; Levin, 1987; Hall, 2006, 2011; Silverman, 2011).² Our description of the observations below will not make strong assumptions about what synchronic status the reported vowels have in their respective phonological system.

While there are claims that vowel insertion is not affected by higher prosodic or intonational factors on the grounds that “phrase-level metrical structures themselves tend to show much optionality” (Hall, 2011:3), there are a number

2 Intrusive vocoids are described as phonetic transitions between consonants that are invisible to the phonological system. Epenthetic vowels are described as phonological segments that are inserted by the grammar (often described as repairing illicit structures, see Hall, 2006).

of languages in which a vowel is more likely to be inserted if there are tune-text conflicts. For example, research on Romance languages frequently mentions vowel insertion patterns conditioned by prosodic and intonational factors.

In one of the more detailed acoustic studies, Grice, Savino, and Roettger (2018) report on experimental evidence from a production study of Bari Italian speakers: In a number of varieties of Italian, the pronunciation of loan words ending in consonants (e.g. *internet*, *blog*) is often characterized by a word final schwa (e.g. Bertinetto, 1985; Krämer, 2009; Repetti, 2012; Broniś, 2016; Cavarani, 2015). In Grice et al.'s study, speakers produced monosyllabic and disyllabic names ending in word-final consonants (e.g. *Matt*, *Tim*, *Carol*), a pattern that is phonotactically rare in their native vocabulary. Targets were produced in different prosodic contexts associated with tonal contours that differed in their nature and complexity. Grice et al. find that both the frequency of occurrence and duration of schwa is modulated by intonational and metrical factors. The probability of schwa occurrence and its duration was robustly affected by the intonational tones to be realized on the word. Rising tones and rise-falls elicit more and longer schwas than falling tones. Moreover, monosyllabic words elicited overall more schwas and if schwa was present, overall longer schwas than disyllabic words. This pattern was more pronounced in cases in which the word co-occurred with a rise-fall.

For example, the word *Carol* in (2) is more likely to be produced with a word-final schwa when the word occurs in a question, characterized by a rise-fall(-rise), than in a statement, characterized by a fall. This tune-dependent pattern was stronger for monosyllabic words than for disyllabic words.

- (2) /a kja'mato 'karɔl(ə)/
 'Carol called' (Grice et al., 2018)

These quantitative assessments are in line with many cross-linguistic observations throughout the literature. For instance, Martínez-Gil (1997) reports on Galician vowel insertion in phrase-final position, a pattern that is very similar to the Bari Italian case. He reports an optional [i] that can occur in phrase-final position. However, he notes that this optional vowel can only occur in monosyllabic words or polysyllabic words with final stress. While he formulates this pattern as categorical, it is strikingly similar to the Italian case in which there are more schwas in monosyllabic words than in disyllabic words. Martínez-Gil attributes these distributional asymmetries to structural requirements to create a well-formed bimoraic trochee at the edge of each intonational phrase. However, Martínez-Gil also states that "sentential prominence in Galician falls on the last word bearing primary stress" and that "epenthesis is heavily favoured

[...] whenever a word bears phrasal prominence" (p. 288), suggesting a correlation between vowel insertion and intonation.

Similarly, in Sardinian, there is word-final vowel insertion which has been described as repairing illicit structures defined by either having a final consonant or final word stress. According to a recent survey in Torres-Tamarit, Linke, and Vanrell (2017), the insertion of an epenthetic vowel does not always take place. Epenthesis is robust in phrase-final position, which is the locus of certain complex tonal movements associated with questions (Vanrell, Ballone, Schirru, & Prieto, 2015), but it is variable in phrase-medial position. For Carrarese and Pontremolese, two Northern varieties of Italian, Cavirani (2015) mentions variable realization of word-final vocoids whose presence and duration are determined by 'emphatic conditioning', which, again, can be interpreted as a prominent (and therefore particularly dynamic) type of pitch accent.

In Standard European Portuguese, vowels are inserted into different prosodic contexts, namely phrase-finally, in polar questions which are characterized by a complex nuclear contour (fall-rise), and in vocative constructions, characterized by a high plateau (Frota, 2002; Frota et al., 2015a). Cruz (2013) compared vowel insertion patterns across different European Portuguese varieties. In Standard European Portuguese, vowel insertion occurs in Intonation Phrase-final position when the word co-occurs with tonal configurations involving a high tone or rise, but not in declaratives in which the word co-occurs with simple falls, whereas in the variety spoken in Castro Verde, vowel insertion takes place in Intonation Phrase-final position across the board, i.e. regardless of the tonal contour.

Similar observations have been made for genetically unrelated languages. For example, in Tashlhiyt Berber inserted vowels interact with phrase-level intonation in systematic ways. Roettger (2017) discusses a production corpus containing lexically voiceless words occurring in different prosodic positions (phrase medially vs. phrase finally) and in different sentence modalities characterized by different intonational events (rise-fall vs. fall vs. no tonal movement). Besides speaker-specific patterns and certain articulatory constraints, the presence of schwa is correlated with both phrasal position and the nature of the tune that is supposed to be realized on the word (see also Grice et al., 2011, 2015). Schwa was more often observed in phrase-final words than in phrase-medial words, and it was more often observed in sentence modalities in which complex tonal movements have to be realized. Fig. 6 illustrates example utterances in which the phrase final word is contrastively focused, characterized by a rise-fall in pitch. If the focused word does not lexically contain tone bearing units (i.e. sonorants, see Roettger, 2017), a non-lexical vowel can be inserted to carry significant portions of the intended tonal configuration (here pitch peak

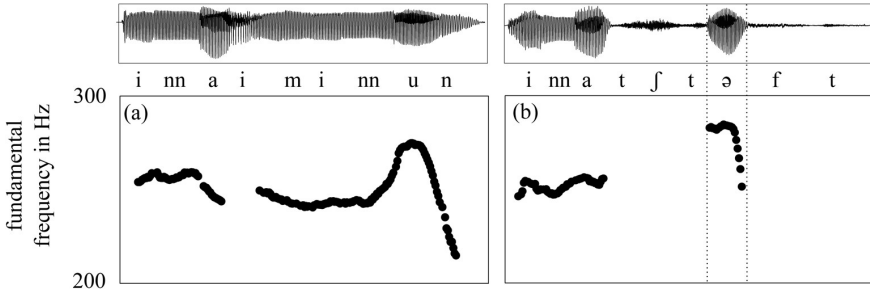


FIGURE 6 Representative waveforms and f_0 contours of the Tashlhiyt Berber utterances (a) “inna iminnun.” ‘He said always.’ (a fully voiced utterance) and (b) “inna tctft.” ‘He said you crushed’ (an utterance containing a phonologically voiceless target word in phrase-final position). The final word in both utterances is in corrective, characterized by a rise-fall in pitch. In (b) the high pitch peak and part of the fall are realized on a non-lexical vowel.

FIGURE ADAPTED FROM ROETTGER (2017: FIGURE 6.5)

and fall). As illustrated in Fig. 6, these tone bearing schwas are characterized by high intensity and long durations.

Moroccan Arabic, a language spoken alongside Tashlhiyt Berber, also has positionally conditioned vowel insertion, formulated as a phonological rule by Dell and Elmedlaoui: “In the last syllable of an Intonational Phrase, if the nucleus does not contain a sonorant, make it complex by inserting *e* before the nuclear consonant.” (2002: 300). Thus, vowel insertion takes place near the end of the intonation phrase. In another variety of Arabic, Tunisian Arabic, schwa is frequently inserted Intonation Phrase-finally in polar questions, where the melody is complex (Hellmuth, in press), but it is not inserted in other sentence modalities. Moreover, vowel insertion is more likely in monosyllabic words and words with final stress, both metrical contexts in which the potential for tonal crowding is maximized.

Cross-linguistically the literature indicates that vowel insertion is more likely to happen in phrase-final position than in phrase-medial position, and it is more likely to happen in environments in which certain communicatively relevant tonal movements take place. Moreover, it appears that there are certain implicational relationships. If vowel insertion is conditioned by tonal crowding, there is either vowel insertion only in those environments, as is the case for Sardinian, Galician, Moroccan Arabic and Tunisian Arabic, or there is at least quantitatively more vowel insertion in those environments, as is the case in Bari Italian. Similarly, if vowel insertion is conditioned by prosodic structure, it is either observed only phrase finally, as is the case for Galician, Moroccan Arabic, and the Portuguese varieties, or there is at least

quantitatively more vowel insertion in phrase-final position, as in Sardinian and Tashlhiyt. Finally, if vowel insertion is conditioned by the tune, it is either restricted to certain contours (mostly contours that contain a rising element) as is the case for Tunisian Arabic and Standard European Portuguese, or there are quantitatively more instances of vowel insertion in those tonal contexts that contain rises in pitch. These observations are in line with the idea that linguistic systems create a segmental environment that is better suited for realizing meaningful f_0 movements. In other words, the text has to be suitable for realizing the tune.

4.3 *Avoiding vowel devoicing / deletion*

Just as certain contexts favor the addition of a vowel to the segmental string, these same sorts of contexts tend to disfavor vowel devoicing or vowel deletion, since vowels (with full voicing) are optimal for the realization of f_0 movements.

Although vowel devoicing is often restricted to high vowels and commonly occurs when the vowel is adjacent to voiceless consonants, prosody can also play a role in determining whether vowels are devoiced or not, although there is much disagreement as to what the relevant prosodic factors might be. Japanese, for example, is described as devoicing high vowels between two voiceless consonants, see example (3). Moreover, several authors have also noted that vowel devoicing is more likely to take place prepausally compared to contexts in phrase-medial position (see example 4) (e.g. Fujimoto, 2015; Kilbourn-Ceron & Sonderegger, 2017, for recent overviews).

(3) /sika/ [s̺ika] 'deer'

(4) /karasu/ [karasu] '(it's a) cow'

These prosodically conditioned devoicing patterns in Japanese are not categorical. For example, utterance-final devoicing is highly variable (e.g. Maekawa & Kikuchi, 2005; Vance, 2008; Kilbourn-Ceron & Sonderegger, 2017). Crucially, vowel devoicing can be “blocked” by a high pitch accent (e.g. Kuriyagawa & Sawashima, 1989; Hirayama, 2009). In their corpus analysis of Japanese spontaneous speech, Kilbourn-Ceron and Sonderegger (2017) found that, all else being equal, a larger phrase boundary following the vowel was correlated with *less* devoicing. This variation in devoicing has been interpreted as conditioned by the presence of boundary tones (e.g. Hirayama, 2009; Fujimoto, 2015).

As in Japanese, languages exhibiting lexical tone or lexical pitch accent tend to block final devoicing when a vowel co-occurs with a high tone (Cheyenne, an Algonquian language: Davis, 1962; Acoma, a Keres language: Miller, 1965; Konso,

an East Cushitic language: Orkaydo, 2013).³ In the Shanghai variety of Wu Chinese, devoicing has been described as never occurring in the first syllable in a tone sandhi group, which is the unit to which the phonological tone is associated (Beckman & Venditti, 2010; and see Chao, 1968, who describes a similar pattern for a variety spoken in Beijing Chinese).

Post-lexical tones show similar patterns in a number of languages. For example, Dell and Tangi (1992) report that in Ath-Sidhar Rifian Berber, schwa is usually devoiced between voiceless consonants but remains voiced “under an intonation which requires a final high pitch [...]” (ibid: 154). Heath (1987: 184) describes a process in Moroccan Colloquial Arabic in which schwa deletion in word-final syllables is blocked by “list Intonation”, characterized by a final rise in pitch. Gordon (1998) reports on vowel devoicing being less likely in accented positions in Bonaara Oromo, and Tunica. Andreeva and Koreman (2003) discuss an interesting case of vowel devoicing in Bulgarian. When a pitch accent occurs on the last word in the phrase (i.e. to signal focus), the amount of vowel devoicing depends on the scaling and alignment of the accent peak, with less devoicing for more prominent accent types. Moreover, they observe no devoicing when the word appears phrase finally and co-occurs with a rising boundary tone. This is in line with observations about vowel devoicing in Greek (Dauer, 1980; Kamaiki, 2015), which is less likely to take place when the target word occurs with rising intonation (as for example in *wh*-questions), rather than falling intonation. In Standard European Portuguese, in words with penultimate stress, word-final schwa deletion may be blocked in intonation-phrase final position under falling-rising contour characteristic of polar questions (Frota et al. 2015b, 2016; Vigário, 2016). This is an interesting case because in this variety of Portuguese, word-final schwa deletion has been described as mandatory, but as regularly failing to apply in intonation phrase-final position in certain tonal contexts.

All these examples strongly suggest a relationship between the application of vowel devoicing / deletion and the tonal conditioning of the environment in which the vowel is to be produced: In positions in which there is a communicatively relevant tonal movement, there is less vowel devoicing. This supports the hypothesis that the tune is to some extent driving the text. Devoiced vowel variants in these positions would lead to the loss of pitch information related to lexical or postlexical meaning, endangering an otherwise important meaning contrast.

3 Several languages also exhibit blocking of vowel devoicing in stressed syllables (see Gordon 1998 for an overview).

5 General discussion

So far we have argued that languages have at their disposal many different ways to resolve the conflict between requirements for pitch movements and the segments that host these pitch movements. This can be achieved by adjustments to the tune, such as truncation, compression, and shifting of tonal targets. These interactions change the shape or timing of relevant tonal movements, adjusting them to accommodate segmental restrictions. This can also be achieved by adjustments to the text, an area that has received less attention to date. In conflicting tune-text settings, languages often lengthen existing vowels or insert non-lexical vowels. Other languages avoid otherwise instantiated vowel devoicing or deletion in such settings. Arguably, all textual adjustments increase the suitability of the text to realize tonal movements by adding or preserving aspects of the text that are optimal for the production and perceptual retrieval of pitch. Focusing on the latter, adjustments to the text, we will now discuss the ways in which observations on segmental adjustments can relate to possible sound change and whether our findings are compatible with recent claims about the evolution of sound systems.

5.1 *Tune-text interactions as phonetically conditioned sound change?*

Assuming that sound systems are shaped by perceptual, articulatory, and acoustic aspects of speech transmission (e.g. Hombert et al., 1979; Ohala, 1981, 1993; Blevins, 2004; see Yu, 2013, for a recent overview), recurring small biases in the channel can function as diachronic attractors, creating certain evolutionary pathways that are more common than others. For example, devoicing of obstruents in domain-final position, a very common sound pattern, is attested in many unrelated languages across the world. This frequently occurring sound pattern might reflect several phonetic tendencies that lead to variation favoring voiceless final obstruents. Smith (1997) suggests that insufficient transglottal airflow might hinder vocal fold vibration in domain-final positions. Steriade (1999) argues that stop releases in final position are often impoverished or missing, thus obscuring important acoustic cues to voicing distinctions. Kohler (2000) suggests that final devoicing might be due to a progressive opening of the glottis for breathing in prepausal position. Blevins (2004) additionally proposes that final lengthening in phrase-final position might lead to a disproportional lengthening of stops endangering the voicing contrast which is often cued by closure duration (i.e. voiced stops have smaller closure durations). Thus, final devoicing demonstrates not only the breadth of phonetic motivations for sound change (aerodynamic, perceptual, articulatory, acoustic), but also the multiplicity of diachronic paths. One and the same outcome

(here a voiced sound becoming voiceless in domain-final position) might result from different properties of the speech transmission process (or a combination thereof). Moreover, this example highlights the role of positional effects for phonetically conditioned sound change (e.g. Cole & Hualde, 2013). All phonetic motivations for final devoicing are either conditional on particular prosodic positions (i.e. occur only in certain prosodic positions) or have a higher probability of occurring in these positions.

The present paper has argued that languages are characterized by phonetic variation that facilitates the production and perceptual retrieval of pitch information. The idea that properties of tonal structures may affect their segmental environment is not new, albeit contested (Hyman, 1976). Some authors have argued that the nature of a lexical tone and its accompanying laryngeal configuration can lead to the diachronic insertion of a segment (e.g. Kock, 1901; Maddieson, 1974, 1976; Gandour & Maddieson, 1976).

In general, there seem to be asymmetries between falling and rising tonal movements on the one hand, and between complex and simple tonal movements on the other. We have discussed cases in which languages adjust the segmental string when co-occurring with tunes that involve a rise, implying that rising tones are special in some way (see Evans, 2015 for discussion of the special status of high tones). Raising and lowering pitch are not achieved in the same way, but instead involve different laryngeal configurations (e.g. Moisik et al., 2014). Producing a rise in pitch generally takes longer than a fall of an equivalent excursion. Moreover, it has been argued that there are physiological limitations on the maximum speed of pitch change (Ohala & Ewan, 1973, Sundberg, 1979, Xu & Sun, 2002), restricting high/rising tones more than low/falling tones. Additionally, there are known correlations between pitch level / pitch movement and perceived segment duration. A vowel with a high-level tone is perceived as longer than a vowel with a low-level tone, and a vowel with a rising tone is perceived as longer than a vowel with a falling tone (Lehiste, 1976; Wang et al., 1976; Yu, Lee, & Lee, 2014). Thus, raised f_0 might be already inherently correlated with added length and larger gestural intervals, in both perception and articulation.

Orthogonal to perceptual biases, increased f_0 has been argued to activate attention orienting (Hsu, Evans, & Lee, 2015). High tones / rising tones may thus drive selective attention, making phrasal positions co-occurring with these pitch dynamics more salient, and thus better targets for reproduction. A small phonetic bias—especially if made salient by pitch prominence—might lead to a drift towards patterns that optimize the transmission of the tune as discussed in this paper.

5.2 *Negotiating between message transmission accuracy and resource cost*

The discussed sources of phonetic variation can potentially be informative for functionalist models of sound systems. It has recently been argued that sound systems are shaped by the trade-off between a bias towards optimizing message transmission accuracy and a bias to reduce resource cost (e.g. Hall et al., 2016, 2018 and references therein).

This framework makes predictions about where in the speech signal it is most beneficial to invest resources to increase transmission accuracy on the one hand, and to keep the message transmission cost-effective on the other. In other words, parts of the signal that contribute more to accuracy are biased toward phonetic enhancement, whereas parts of the signal that contribute less to accuracy are biased toward phonetic reduction. It is important to note that the proposed bias is argued to be only one of many biases that influence sound systems.

This approach subsumes a variety of different findings throughout the literature on sound patterns (see Hall et al., 2016 for an overview). For example, phonological contrasts with high functional load (i.e. contrasts that distinguish many minimal pairs) are less likely to undergo neutralizations (as an instance of resource cost reduction) than contrasts with low functional load (Wedel, Kaplan, & Jackson, 2013). For instance, in a corpus of 50 different languages, Wedel et al. (2018) found that neutralizations are statistically more common at the end of words than at the beginning of words. Given that listeners recognize words incrementally as the phonetic signal unfolds (e.g. Allopenna, Magnuson, & Tanenhaus, 1998; Dahan et al., 2001), it could be argued that word-final position contributes only little to word recognition and thus neutralizations in this position do not compromise message transmission accuracy.

These approaches, however, have mainly focused on aspects of speech that signal lexical meaning and do not capture other important meaning signaling dimensions of speech, including prosodic and social dimensions (see also Foulkes et al., 2018). The present paper has reviewed a body of research that suggests an intricate interaction between these different meaning signaling layers of speech, i.e. the segments signaling lexical meaning and the intonation signaling pragmatic meaning.

Some phenomena increase resource costs without reducing message transmission accuracy in terms of word recognition and, at the same time, facilitate the retrieval of pragmatic meaning or important cues to phrasal structure. For example, adding a non-lexical vowel or inhibiting the reduction of an otherwise present vowel in phrase-final position might alter the segmental make-up of a word. In light of its prosodic position, this segmental adjustment does not

reduce the message transmission accuracy of the word (because the word is already highly predictable from its context; the intended word can be retrieved based on the beginning of the word, Hall et al., 2016). However, the addition / preservation of segmental material with high periodic energy and rich harmonic structure might facilitate the realization and retrieval of pitch movements which function as important cues to discourse-relevant functions.

The patterns in the literature suggest that we need to take the transmission accuracy of non-lexical aspects of the message into account in order to fully understand how sound systems allocate resources. The *tune transmission accuracy*, i.e. the likelihood of retrieving intonational meaning correctly, is achieved through a text that is optimized for pitch contours to be articulated and perceived. This is in line with findings that suggest that the interpretation of intonational contours is strongly impaired in the context of impoverished text (short vowels, voiceless consonants etc.) resulting in a truncation of the tune (Odé, 2005; Rathcke, 2013).⁴

Previous work on the optimization of sound systems for message transmission accuracy versus resource cost has focused on transmission of lexical meanings. The findings we report here show that the message carried by the tune is also part of this global optimization. All of the phenomena described here (insertion of vowels, avoiding vowel devoicing/deletion, vowel lengthening) increase the resource cost from a lexical perspective (i.e., more segments to articulate, distortion of the form of a word), but they also significantly increase the tune transmission accuracy. It is thus clear that the requirements of tune and text, both in accuracy and resource cost, would have to be negotiated. This negotiation might affect how sound systems are shaped through time.

In principle, this account might be a promising departure point for explaining why some phonological systems favor tune adjustments over text adjustments (Frota et al. 2015b, 2016). Theoretically, message transmission accuracy could be quantified for all relevant levels of the ‘message’ including pragmatic and lexical meaning so as to derive the relative gain / loss for different adjustment strategies. If a text adjustment seriously impairs the retrieval of lexical information, i.e. inserting a non-lexical vowel produces a surface form which resembles another word or changes the stress pattern of the word, a tune adjustment (which might reduce the transmission accuracy of intonational meaning) might be overall better optimized for the system. In contrast, a system that conveys important pragmatic intentions by local intonational events

4 Albeit arguably reducing resource costs, truncation, compression or tonal shift (see section 3) are decreasing the tune transmission accuracy by reducing (truncation), compressing (compression), or moving information away from its locus of interpretation (shift).

only (i.e. in the absence of morphosyntactic devices), adjustments of the text, even at a cost of reduced message transmission accuracy for lexical items, might be preferred.⁵

5.3 *Limitations of this account*

There are two aspects of our account that warrant further explanation. First, if the driving force behind tune-text interactions is communicative in nature (i.e. increasing message transmission accuracy), why do we overwhelmingly find the adjustments of the text when the tune contains a rise? If the motivations for the described segmental changes are solely communicative, there would also be cases where falling pitch movements or low pitch targets require enhancement and should therefore be subject to the same effects. One might argue that while rises often mark phrase boundaries, signal continuation and mark questions and contrastivity, falling pitch movements or flat stretches of pitch seldom trigger these segmental changes because, in many languages, they signal the default interpretation, indicating unmarked declaratives (Lieberman 1967; Gussenhoven 2016). In Hall et al.'s model (2016, 2018), this default would constitute a prior, i.e. the most likely interpretation in the absence of additional information. Only if the speakers need to signal a divergence from the prior do they invest extra resources to uphold message transmission accuracy.⁶ Alternatively, one could speculate that the attentional asymmetries between high and low pitch discussed above make the learners' generalizations of such patterns more likely for tonal configurations with a high pitch component.

Second, there is so far only sparse evidence for categorical segmental alternations or sound change. Why is it that we do not see more phonologized instances of the proposed interactions? We have discussed many different

5 The proposed interaction between functional load and text-tune adjustment is of course a simplification and will constitute only a weak bias in the evolution of sound systems. For example, there are Italian varieties that truncate instead of inserting a vowel. It might be speculated that global or distal prosodic cues in the utterance are sufficient for conveying a pragmatic contrast, allowing local intonational events to be truncated without endangering message transmission accuracy.

6 Note that there are cases in which segmental changes are not triggered by tonal events with a rising component. For example, Cruz (2013) discusses the southern European Portuguese dialect spoken in Alentejo, in which vowel insertion also frequently occurs in declaratives, characterized by monotonal contour. In fact, many of the empirical studies discussed in this paper show segmental adjustments like vowel insertion for falling contours, however, in most cases these relationships are probabilistically less common. We are left with acknowledging that there have to be other sources for domain-final vowel insertion that might be orthogonal to the proposed functional explanation in this paper.

cases of segmental adjustments that are conditioned by the tunes they co-occur with. Most of these observations suggest a probabilistic relationship, i.e. conditional on a particular tune, a segmental adjustment such as vowel insertion or vowel devoicing is more or less likely. Only a few adjustments are actually described as categorical or rule-based: The Intonation Phrase final schwa epenthesis in Moroccan Arabic was described by Dell and Elmedlaoui (2002) as a phonological rule; Intonation Phrase final blocking of word-final vowel deletion in Portuguese described by Frota et al. (2015b) also suggests a categorical pattern. Despite these cases, the obvious question arises as to why the in-the-moment changes to the texts that were described in this paper do not lead more frequently to long-term systematic changes in phonological systems. We offer several speculative answers to this question:

First, there might be more cases in which such adjustments have phonologized, but have not been reported. Phrase-level prosodic patterns are still a heavily underdocumented area of linguistic descriptions and we lack historic records of these aspects of sound systems. Thus, segmental changes conditioned by phrase-level prosody might have been overlooked or discarded.

Alternatively, one could argue that the specific conditions under which these segmental adjustments happen are too rare for language users to generalize. We have described several examples of changes to the text that were conditional on specific pragmatic contexts (and their accompanying tune requirements). For example, in polar questions, speakers of Tashlhiyt (Roettger, 2017) or Tunisian Arabic (Hellmuth, in press) produced more inserted non-lexical vowels than in other sentence modalities. Language users might not encounter these cases frequently enough to generalize them. Alternatively, because of their specific context, these phenomena might not be interpreted phonologically, i.e. as a phonological alternation, but rather as morphological or syntactic in nature. Vocative constructions are a case in point. Vocative constructions are used to call out to interlocutors or attract / maintain addressees' attention (e.g. Daniel & Spencer, 2009) and they are often described as being grammatically marked by affixes or syntactic particles. These constructions have also often been described as exhibiting a specific intonational contour, referred to as vocative chant (Lieberman, 1975), stylized fall (Ladd, 1978), or chanted call (Hayes & Lahiri, 1991). These tunes usually consist of a rising element followed by a sustained mid to high plateau or fall, a tune that necessitates sufficient text to be realized. There are several descriptions of vocative constructions that exhibit phonological alternations akin to what has been discussed in the present paper: For example, in Kannada, a Dravidian language (Schiffman, 1983), and Eastern Burushaski, an isolate spoken in Pakistan, the stem-final vowel is lengthened in vocatives (Noboru, 2012); In Aguaruna and Wampis,

both Chicam languages, the vocative is characterized by blocking of otherwise regular deletion of the final vowel (Overall, 2007; Pena, 2016); In Chukchi, a Chukotko–Kamchatkan language, epenthetic schwas and reduced vowels are promoted to full vowels (Dunn, 1999). These alternations are described in terms of morphological processes akin to affixation or stem alternation.

Beyond morphology, language users might re-analyze segmental changes driven by the tune as syntactic particles. Hellmuth's work (in press) on vowel insertion in Tunisian Arabic shows that speakers insert vowels in phrase-final position only in polar questions. There are virtually no instances of phrase final vowel insertions in other sentence modalities. Hellmuth speculates that this instance of vowel insertion can be conceived of as a question particle. If within a linguistic system, certain pragmatic functions systematically co-occur with certain tonal configurations in salient prosodic positions, and the realization of those configurations is facilitated by changes to the text such as vowel insertion, frequently co-occurring changes to the text might be interpreted by the language learner as discourse markers rather than as phonological alternations. Instead of generalizing the pattern across the lexicon, they might generalize them across specific discourse-pragmatic contexts.

German and Prévot (2014) show that in Singapore English, the sentence-final particle *lah*, is affected by its prosodic position. *Lah* which has been described as fulfilling social and pragmatic functions (Besemeres & Wierzbicka, 2003), is more likely to occur after a lexically stressed syllable than after an unstressed syllable. German and Prévot argue that the particle is recruited to avoid tonal crowding in the form of an intonational rise-fall on the final stressed syllable of the utterance. This observation also resonates well with the adjustment to the text discussed in this paper, as the text (in this case the particle) is selected according to the needs of the tune.

These findings open up the possibility that tune-text interactions potentially find their way into the grammatical system without triggering regular sound change. Thus, an awareness of how the text may adjust to the tune in order to enable communicatively relevant intonational cues may further our understanding of the evolution of grammar more generally.

6 Conclusion

Several scholars have speculated on the origin of speech: Charles Darwin wrote about our early ancestors using singing to mediate sexual and territorial issues. He referred to this as a musical proto language (Darwin, 1871). Darwin's ideas have been adopted and expanded by several scholars, who proposed similar

proto languages (Brown, 2000; Fitch, 2005; Mithen, 2005), all of which assign a central role to rhythm and melody in the evolution of human speech. This paper has argued that grammatical aspects of rhythm and melody—what we refer to as prosodic structure and intonation—may play an important role in shaping phonological systems, too.

The requirements for the expression of pragmatic and lexical meanings sometimes conflict with each other, calling for a negotiation between tune and text. In this paper, we have argued that intonation places functional pressure on its segmental hosts which can facilitate the temporal adjustment and preservation of existing elements—or the insertion of new elements—to increase transmission accuracy at the pragmatic level. By distilling previously unconnected findings across languages and linking them to phonetic aspects of the speech transmission process, we aim to understand why certain phonological processes and alternations observed across languages happen. More broadly, we invite the field to further explore why phonological systems are the way they are. We believe that taking the intricate relationship between tune and text seriously will not only advance our knowledge of intonation systems, but it will also add a new dimension to the question as to how the message shapes the phonological form.

Acknowledgments

Funding from the German Research Foundation (SFB 1252 Prominence in Language) is gratefully acknowledged. We would like to thank Aviad Albert, James German, Márton Sóskuthy, and Andy Wedel for their comments on earlier drafts of this manuscript. We would like to thank the Departments of Linguistics at the University of Cologne, the Goethe University in Frankfurt, Northwestern University in Evanston, the Institute of Phonetics and Speech Processing at the University of Munich, and the audience of XlanS in Lyon for their valuable input on this project. Finally, we would like to express our gratitude to three anonymous reviewers and the editor Jeff Good for their comments during the review process. All remaining errors are our own.

References

- Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Alloppenna, Paul D., James S. Magnuson, & Michael K. Tanenhaus. 1998. Tracking the

- time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38(4): 419–439.
- Andreeva, Bistra & Jacques Koreman. 2003. The status of vowel devoicing in Bulgarian: Phonetic or phonological? In Gerhild Zybatow, Luka Szucsich, Uwe Junghanns & Roland Meyer (Eds.), *Formal Description of Slavic Languages: The Fifth Conference (FDSL5)*, Leipzig, 81–91. Frankfurt: Peter Lang.
- Bannert, Robert & Anne-Christine Bredvad-Jensen. 1975. Temporal organisation of Swedish tonal accent: the effect of vowel duration. *Working Papers in Linguistics, Lund University* 10: 1–36.
- Barnes, Jonathan, Alejna Brugos, Nanette Veilleux, & Stefanie Shattuck-Hufnagel. 2014. Segmental influences on the perception of pitch accent scaling in English. *Proceedings of Speech Prosody*, 7:1125–1129.
- Beckman, Mary E. & Jennifer J. Venditti. 2010. Tone and intonation. In William G. Hardcastle, John Laver, & Fiona E. Gibbon (Eds.), *The Handbook of Phonetic Sciences*, 2nd ed., 603–652. New Jersey: John Wiley & Son.
- Bertinetto, Pier Marco. 1985. A proposito di alcuni recenti contributi alla prosodia dell'italiano. *Annali Della Scuola Normale Superiore Di Pisa. Classe Di Lettere e Filosofia* 15(2): 581–643.
- Besemeres, Mary & Wierzbicka, Anna. 2003. The meaning of the particle lah in Singapore English. *Pragmatics & Cognition* 11(1), 3–38.
- Biondi, Roberto Zariquiey. 2011. *A grammar of Kashibo-Kakataibo*. PhD Dissertation, La Trobe University.
- Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Bolinger, Dwight Le Merton. 1965. Pitch Accent and Sentence Rhythm. In Isamu Abe & Tetsuya Kanekiyo (Eds.), *Forms of English: Accent, Morpheme, Order*, 139–180. Tokyo: Hokuou.
- Bradshaw, Mary M. 1999. *A crosslinguistic study of consonant-tone interaction*. PhD Dissertation, The Ohio State University.
- Broniś, Olga. 2016. Italian vowel paragoge in loanword adaptation. Phonological analysis of the Roman variety of Standard Italian. *Italian Journal of Linguistics* 28(2): 25–68.
- Brown, Steven. 2000. The 'Musilanguage' Model of Music Evolution. In Nils Lennart Wallin, Björn Merker, & Steven Brown (Eds.), *The Origins of Music*, 271–300. Cambridge, Mass: The MIT Press.
- Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. PhD Dissertation, Lund University.
- Cabré, Teresa & Maria del Mar Vanrell. 2013. Entonació i truncament en els vocatius romànics. *Actes del 26è Congrés de Lingüística i Filologia Romànica*, Vol. 1, 543–553. Berlin: de Gruyter.

- Cangemi, Francesco & Martine Grice. 2016. The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1): 1–20.
- Caspers, Johanneke & Vincent J. van Heuven. 1993. Effects of time pressure on the phonetic realization of Dutch accent-lending pitch rise and fall. *Phonetica* 50(3): 161–171.
- Cavirani, Edoardo. 2015. *Modeling Phonologization: Vowel Reduction and Epenthesis in Lunigiana Dialects*. LOT: Utrecht.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Cole, Jennifer S. & José I. Hualde. 2013. Prosodic structure in sound change. In Shu-Fen Chen & Benjamin Slade (Eds.) *Grammatica Et Verba/Glamor and Verve: Studies in South Asian, Historical, and Indo-European Linguistics in Honor of Hans Henrich Hock on the Occasion of His Seventy-Fifth Birthday*, 28–45. Ann Arbor, MI: Beech Stave Press.
- Cruz, Marisa. 2013. *Prosodic variation in European Portuguese: Phrasing, intonation and rhythm in central-southern varieties*. PhD Dissertation, Universidade de Lisboa.
- Crystal, David. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.
- Dahan, Delphine, James S. Magnuson, Michael K. Tanenhaus, & Ellen M. Hogan. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16(5–6): 507–534.
- Daniel, Michael & Andrew Spencer. 2009. The vocative—an outlier case. In Andrej L. Malchukov & Andrew Spencer (Eds.), *The Oxford Handbook of Case*, 626–634. Oxford: Oxford University Press.
- Darwin, Charles. 1871. *The Descent of Man and Selection in Relation to Sex*. London: John Murray.
- Dauer, Rebecca M. 1980. The reduction of unstressed high vowels in Modern Greek. *Journal of the International Phonetic Association* 10(1–2): 17–27.
- Davis, Irvine. 1962. Phonological Function in Cheyenne. *International Journal of American Linguistics* 28 (1): 36–42.
- Dell, François & Mohamed Elmedlaoui. 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer.
- Dell, François & Oufae Tangi. 1992. Syllabification and empty nuclei in Ath-Sidhar Rifian Berber. *Journal of African Languages and Linguistics* 13(2): 125–162.
- D'Imperio, Mariapaola. 2001. Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33(4): 339–356.
- D'Imperio, Mariapaola & David House. 1997. Perception of questions and statements in Neapolitan Italian. In *Proceedings of the 5th European Conference on Speech Communication and Technology*, 251–254. Grenoble: ESCA.

- Dunn, Michael John. 1999. *A grammar of Chukchi*. PhD Dissertation, Australian National University.
- Erikson, Y. & Margit Alstermark. 1972. Fundamental frequency correlates of the grave word accent in Swedish: The effect of vowel duration. *Speech Transmission Laboratory, Quarterly Papers and Status Report*, 2–3: 53–60.
- Evans, Jonathan Paul. 2015. High is not just the opposite of low. *Journal of Phonetics* 51: 1–5.
- Fitch, W. Tecumseh. 2005. Protomusic and protolanguage as alternatives to Protosign. *Behavioral and Brain Sciences* 28(2): 132–133.
- Foulkes, Paul, Gerry Docherty, Stefanie Shattuck-Hufnagel, & Vincent Hughes. 2018. Three steps forward for predictability. Consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard* 4 (s2).
- Frota, Sónia. 2002. Tonal association and target alignment in European Portuguese nuclear falls. In Natasha Warner & Carlos Gussenhoven (Eds.), *Papers in Laboratory Phonology* 7, Vol. 7, 387–418. Berlin: Mouton de Gruyter.
- Frota, Sónia. 2014. The intonational phonology of European Portuguese. In Sun-Ah Jun (ed.), *Prosodic Typology 11: The Phonology of Intonation and Phrasing*, 6–42. Oxford: Oxford University Press.
- Frota, Sónia, Joelma Castelo, Marisa Cruz, Verónica Crespo-Sendra, Nádia Barros, Aline Silvestre, & Marina Vigário. 2015a. Melodia ou texto? Estratégias de acomodação entre melodia e texto no Português. *Diadorim: Revista de Estudos Linguísticos e Literários* 17(2), 12–33.
- Frota, Sónia, Marisa Cruz, Flaviane Fernandes-Svartman, Gisela Collischonn, Aline Fonseca, Carolina Serra, Pedro Oliveira, & Marina Vigário. 2015b. Intonational variation in Portuguese: European and Brazilian varieties. In Sónia Frota and Pilar Prieto (Eds.), *Intonational Variation in Romance*, 235–283. Oxford: Oxford University Press.
- Frota, Sónia, Marisa Cruz, Joelma Castelo, Nádia Barros, Verónica Crespo-Sendra, & Marina Vigário. 2016. Tune or text? Tune-text accommodation strategies in Portuguese. In Jon Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel, & Nanette Veilleux (Eds.), *Proceedings of Speech Prosody* 8, 692–696. Boston: ISCA.
- Fujimoto, Masako. 2015. Vowel devoicing. In Haruo Kubozono (ed.), *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, 167–214. Berlin: Mouton de Gruyter.
- Gandour, Jack & Ian Maddieson. 1976. Measuring larynx movement in Standard Thai using the cricothyrometer. *Phonetica* 33(4): 241–267.
- Gartenberg, Robert & Christa Panzlaff-Reuter. 1991. Production and perception of F₀ peak patterns in German. *Arbeitsberichte Des Instituts für Phonetik der Universität Kiel* 25: 29–115.

- German, James Sneed & Laurent Prévot. (2014). Phonology constrains the distribution of the particle *lah* in Singapore English. Abstract at *LabPhon* 14, 25–27 July, Tokyo.
- Gilles, Peter. 2005. *Regionale Prosodie Im Deutschen: Variabilität in Der Intonation von Abschluss Und Weiterweisung*. Berlin: Walter de Gruyter.
- Goldsmith, John A. 1974. An autosegmental typology of tone: And how Japanese fits in. In Ellen Kaisse & Jorge Hankamer (Eds.), *Proceedings from the Fifth Regional Meeting of the North East Linguistics Society*, 172–182. Cambridge MA: Harvard University Linguistics Department.
- Goldsmith, John A. 1976. *Autosegmental Phonology*. Bloomington, IN: Indiana University Linguistics Club.
- Gordon, Matthew. 1998. The phonetics and phonology of non-modal vowels: A cross-linguistic perspective. *Annual Meeting of the Berkeley Linguistics Society* 24: 93–105.
- Grabe, Esther. 1998. Pitch accent realization in English and German. *Journal of Phonetics* 26(2): 129–143.
- Grabe, Esther, Brechtje Post, Francis Nolan, & Kimberley Farrar. 2000. Pitch accent realization in four varieties of British English. *Journal of Phonetics* 28(2): 161–185.
- Grice, Martine. 1995. *The Intonation of Interrogation in Palermo Italian: Implications for Intonation Theory*. Tübingen: Niemeyer.
- Grice, Martine, Rachid Ridouane, & Timo B. Roettger. 2015. Tonal association in Tashlihyt Berber: Evidence from polar questions and contrastive statements. *Phonology* 32(2): 241–266.
- Grice, Martine, Timo B. Roettger, Rachid Ridouane, & Cécile Fougeron. 2011. The association of tones in Tashlihyt Berber. In Wai-Sum Lee & Eric Zee (Eds.) *Proceedings of the 17th International Congress of Phonetic Sciences*, 775–778. Hong Kong: City University of Hong Kong.
- Grice, Martine, Michelina Savino, & Mario Refice. 1997. The intonation of questions in Bari Italian: Do speakers replicate their spontaneous speech when reading. *Phonus* 3: 1–7.
- Grice, Martine, Michelina Savino, & Timo B. Roettger. 2018. Word final schwa is driven by intonation—the case of Bari Italian. *The Journal of the Acoustical Society of America* 143(4): 2474–2486.
- Grønnum, Nina. 1989. Stress group patterns, sentence accents and sentence intonation in Southern Jutland (Sonderborg and Tønder)—with a view to German. *Annual Reports of the Institute of Phonetics, University of Copenhagen* 23: 1–85.
- Gussenhoven, Carlos. 2016. Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science* 8: 425–434.
- Hall, Kathleen Currie, Elizabeth Hume, Florian Jaeger, & Andrew Wedel. 2016. The message shapes phonology. Unpublished manuscript.
- Hall, Kathleen Currie, Elizabeth Hume, Florian Jaeger, & Andrew Wedel. 2018. The role of predictability in shaping phonological patterns. *Linguistics Vanguard* 4(s2).

- Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23(03): 387–429.
- Hall, Nancy. 2011. Vowel epenthesis. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume, & Keren Rice (Eds.), *The Blackwell Companion to Phonology*, 1576–1596. Oxford: Wiley-Blackwell.
- Halliday, Michael Alexander Kirkwood. 1967. *Intonation and Grammar in British English*. The Hague: Mouton de Gruyter.
- Hanssen, Judith. 2017. *Regional Variation in the Realization of Intonation Contours in the Netherlands*. Utrecht: LOT.
- Hanssen, Judith, Jörg Peters, & Carlos Gussenhoven. 2007. Phrase-Final Pitch Accommodation Effects in Dutch. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of 16th International Congress of Phonetic Sciences*, 1077–1080. Universität des Saarlandes: Saarbrücken.
- Harms, Robert T. 1976. The segmentalization of Finnish ‘nonrules.’ *Texas Linguistic Forum* 5: 73–88.
- Haudricourt, André-Georges. 1954. De l’origine des tonse en vietnamien. *Journal Asiatique* 242: 69–82
- Hayes, Bruce and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory* 9(1): 47–96.
- Heath, Jeffrey. 1987. *Ablaut and Ambiguity: Phonology of a Moroccan Arabic Dialect*. Albany, NY: SUNY Press.
- Hellmuth, Sam. in press. Text-tune alignment in Tunisian Arabic Yes-No questions. In Marisa Cruz, Sónia Frota, & Pedro Oliveira (Eds.), *Prosodic Variation (with)in Languages: Intonation, Phrasing and Segments* (Studies in Phonetics and Phonology). Sheffield: Equinox.
- Heston, Tyler M. 2014. Prosodic differences between declaratives and polar questions in Fataluku. In *28th Pacific Asia Conference on Language, Information and Computation*, 395–403.
- Hirayama, Manami. 2009. *Postlexical prosodic structure and vowel devoicing in Japanese*. PhD Dissertation, University of Toronto.
- Hombert, Jean-Marie, John J. Ohala, & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55: 37–58.
- Hsu, Chun-Hsien, Jonathan P. Evans, & Chia-Ying Lee. 2015. Brain responses to spoken F₀ changes: Is H special? *Journal of Phonetics* 51: 82–92.
- Hyman, Larry M. 1973. The Role of Consonant Types in Natural Tonal Assimilations. *Consonant Types and Tone, Southern California Occasional Papers in Linguistics* 1: 149–179.
- Hyman, Larry M. 1976. On some controversial questions in the study of consonant types and tone. *UCLA Working Papers in Phonetics: Studies on Perception and Production of Tone* 33: 90–98.

- Kahn, D. 1976. *Syllable-based Generalizations in English Phonology*. Bloomington, IN: University of Indiana Linguistics Club.
- Kaimaki, Marianna. 2015. Voiceless Greek vowels. The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow.
- Kilbourn-Ceron, Oriana & Morgan Sonderegger. 2017. Boundary phenomena and variability in Japanese high vowel devoicing. *Natural Language & Linguistic Theory* 36(1): 1–43.
- Kock, Axel. 1901. *Die Alt-Und Neuschwedische Accentuierung*. Straßburg: Karl J. Trübner.
- Kohler, Klaus J. 2000. Linguistic and paralinguistic functions of non-modal voice in connected speech. *Proceedings of the 5th Seminar on Speech Production: Models and Data*, Kloster Seeon, Germany, 89–92.
- Krämer, Martin. 2009. *The Phonology of Italian*. Oxford: Oxford University Press.
- Kuriyagawa, Fukuko & Masayuki Sawashima. 1989. Word accent, devoicing and duration of vowels in Japanese. *Annual Bulletin of the Research Institute of Language Processing* 23: 85–108.
- Ladd, D. Robert. 1978. Stylized intonation. *Language* 54(3): 517–540.
- Ladd, D. Robert. 2008. *Intonational Phonology*. 2nd edition. Cambridge: Cambridge University Press.
- Leben, William R. 1973. *Suprasegmental phonology*. PhD Dissertation, Massachusetts Institute of Technology.
- Leben, William R. 1976. The tones in English intonation. *Linguistic Analysis* 2(1): 69–107.
- Lehiste, Ilse. 1976. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics* 4(2): 113–117.
- Levin, Juliette. 1987. Between Epenthetic and Excrescent Vowels. *Proceedings of the 6th West Coast Conference on Formal Linguistics* 6: 187–202. University of Arizona.
- Liberman, Mark Yoffe. 1975. *The intonational system of English*. PhD Dissertation, Massachusetts Institute of Technology, Cambridge.
- Lieberman, Philip. 1967. *Intonation, Perception and Language*. Cambridge, MA: MIT Press.
- Lickley, Robin J., Astrid Schepman, & D. Robert Ladd. 2005. Alignment of 'Phrase Accent' lows in Dutch falling rising questions: Theoretical and methodological implications. *Language and Speech* 48(2): 157–183.
- Maddieson, Ian. 1974. A note on tone and consonants. *UCLA Working Papers in Phonetics*, 27: 18–27.
- Maddieson, Ian. 1976. A further note on tone and consonants. *UCLA Working Papers in Phonetics* 33: 131–159.
- Maekawa, Kikuo & Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In Jeroen van de Weijer, Kensuke Nanjo, & Tetsuo Nishihara (Eds.), *Voicing in Japanese*, 84: 205–228. Berlin: De Gruyter.

- Martínez-Gil, Fernando. 1997. Word-final epenthesis in Galician. In Fernando Martínez-Gil & Alfonso Morales-Front (Eds.), *Issues in the Phonology and Morphology of the Major Iberian Languages*, 270–340. Washington, DC: Georgetown University Press.
- Miller, Wick R. 1965. *Acoma Grammar and Texts*. Berkeley: University of California Press.
- Mithen, Steven. 2005. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld & Nicholson.
- Moisik, Scott R., Hua Lin, & John H. Esling. 2014. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association* 44(1): 21–58.
- Mücke, Doris, Martine Grice, Johannes Becker, & Anne Hermes. 2009. Sources of variation in tonal alignment. Evidence from acoustic and kinematic data. *Journal of Phonetics* 37(3): 321–338.
- Noboru, Yoshioka. 2012. *A reference grammar of Eastern Burushaski*. PhD Dissertation, Tokyo University of Foreign Studies.
- Odé, Cecilia. 2005. Neutralization or truncation? The perception of two Russian pitch accents on utterance-final syllables. *Speech Communication* 47(1–2): 71–79.
- Ohala, John J. 1981. Articulatory constraints on the cognitive representation of speech. *Advances in Psychology* 7: 111–122.
- Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*, 237–278. London: Longman.
- Ohala, John J. & William G. Ewan. 1973. Speed of pitch change. *The Journal of the Acoustical Society of America* 53(1): 345–345.
- Orkaydo, Ongaye Oda. 2013. *A Grammar of Konso*. Utrecht: LOT.
- Overall, Simon E. 2007. *A grammar of Aguaruna*. PhD Dissertation, La Trobe University.
- Peña, Jaime. 2016. *A grammar of Wampis*. PhD Dissertation, University of Oregon.
- Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. PhD Dissertation, Massachusetts Institute of Technology.
- Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.
- Prieto, Pilar, Mariapaola D'Imperio, & Barbara Gili Fivela. 2005. Pitch accent alignment in Romance: Primary and secondary associations with metrical structure. *Language and Speech* 48(4): 359–396.
- Prieto, Pilar & Marta Ortega-Llebaria. 2009. Do complex pitch gestures induce syllable lengthening in Catalan and Spanish. In Marina Vigário, Sónia Frota, & Maria J. Freitas (Eds.), *Phonetics and Phonology: Interactions and Interrelations*, 51–70. Amsterdam: John Benjamins.
- Prieto, Pilar, Jan van Santen, & Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23: 429–451.

- Rathcke, Tamara. 2009. *Komparative Phonetik und Phonologie der Intonationssysteme des Deutschen und Russischen*. Vol. 29. München: Herbert Utz Verlag.
- Rathcke, Tamara. 2013. On the neutralizing status of truncation in intonation: A perception study of boundary tones in German and Russian. *Journal of Phonetics* 41(3–4): 172–185.
- Refice, Mario, Michelina Savino, & Martine Grice. 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*, 783–786. Grenoble: ESCA.
- Repetti, Lori. 2012. Consonant-final loanwords and epenthetic vowels in Italian. *Catalan Journal of Linguistics* 11: 167–188.
- Roettger, Timo B. 2017. *Tonal Placement in Tashlhiyt: How an Intonation System Accommodates to Adverse Phonological Environments*. Berlin: Language Science Press.
- Schepman, Astrid, Robin Lickley, & D. Robert Ladd. 2006. Effects of vowel length and ‘Right Context’ on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34 (1): 1–28.
- Schiffman, Harold. 1983. *A Reference Grammar of Spoken Kannada*. Seattle: University of Washington Press.
- Selkirk, Elisabeth O. 1984. *Phonology and Syntax: The Relationship between Sound and Structure*. Cambridge, MA: MIT Press.
- Shattuck-Hufnagel, Stefanie, Mari Ostendorf, & Ken Ross. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22(4): 357–388.
- Silverman, Daniel. 2011. Schwa. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume, & Keren Rice (Eds.), *The Blackwell Companion to Phonology*, 628–642. Oxford: Wiley-Blackwell.
- Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25(4): 471–500.
- Steele, Shirley A. 1986. Nuclear accent F₀ peak location: Effects of rate, vowel, and number of following syllables. *The Journal of the Acoustical Society of America* 80(S1): S51–S51.
- Steriade, Donca. 1999. Phonetics in phonology: The case of laryngeal neutralization. In Matthew Gordon (Ed.), *Papers in Phonology, Vol. 3, UCLA Working Papers in Linguistics* 2, 25–145. Los Angeles: Department of Linguistics, University of California.
- Strunk, Klaus. 1983. *Typische Merkmale von Fragesätzen Und Die Altindische “Pluti.”* Munich: Verlag der Bayerischen Akademie der Wissenschaften.
- Sundberg, Johan. 1979. Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7(2): 71–79.
- Tang, Katrina Elizabeth. 2008. *The Phonology and Phonetics of Consonant-Tone Interaction*. Los Angeles: University of California Press.

- Tench, Paul. 1996. Intonation and the differentiation of syntactic patterns in English and German. *International Journal of Applied Linguistics* 6(2): 223–256.
- Tilsen, Sam. 2012. Utterance preparation and stress clash: Planning prosodic alternations. In Susanne Fuchs, Pascal Perrier, Melanie Weirich, & Daniel Pape (Eds.), *Speech Production and Perception: Planning and Dynamics*, 119–156. Frankfurt: Peter Lang.
- Torres-Tamarit, Francesc, Kathrin Linke, & Maria del Mar Vanrell. 2017. Opacity in Campidanian Sardinian metaphony. *Natural Language & Linguistic Theory* 35(2): 549–576.
- Trager, George L. & Henry Lee Smith. 1951. *An Outline of English Structure*. Norman, OK: Battenberg Press.
- Vance, Timothy J. 2008. *The Sounds of Japanese*. Cambridge: Cambridge University Press.
- Vanrell, Maria del Mar, Francesc Ballone, Carlo Schirru, & Pilar Prieto. 2015. Sardinian intonational phonology: Logudorese and Campidanese varieties. In Sónia Frota & Pilar Prieto (Eds.), *Intonational Variation in Romance*, 317–349. Oxford: Oxford University Press.
- Vanrell, Maria del Mar & Teresa Cabré. 2011. Troncamento e Intonazione Dei Vocativi in Italia Centromeridionale. *Contesto Comunicativo e Variabilità Nella Produzione e Percezione Della Lingua, Atti El 7 Convegno AISV (26–28 January 2011, Lecce)*. Roma: Bulzoni.
- Vigário, Marina. 2016. Segmental phenomena and their interactions: Evidence for prosodic organization and the architecture of grammar. In Susann Fischer & Christoph Gabriel (Eds.), *Manual of Grammatical Interfaces in Romance*, 41–73. *Manuals of Romance Linguistics*. Berlin: de Gruyter.
- Wang, William S.Y., Ilse Lehiste, Chin-Kuang Chuang, & Nancy Darnovsky. 1976. Perception of vowel duration. *The Journal of the Acoustical Society of America* 60(S1): S92–S92.
- Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61(4): 747–776.
- Wedel, Andrew, Abby Kaplan, & Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2): 179–186.
- Wedel, Andy, Jonathon Geary, Jaycie Martin, Adam Ussishkin, & Adam King. 2018. Phonological grammars evolve to preserve information at word beginnings. In C. Cuskley, M. Flaherty, H. Little, Luke McCrohon, A. Ravignani, & T. Verhoeft (Eds.), *The Evolution of Language: Proceedings of the 12th International Conference (EVOLANGXII)*, 525–527. Toruń, Poland: NCU Press.
- Xu, Yi and Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America* 111(3): 1399–1413.
- Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

- Yu, Alan C.L. 2013. *Origins of Sound Change: Approaches to Phonologization*. Oxford: Oxford University Press.
- Yu, Alan C.L., Hyunjung Lee, & Jackson Lee. 2014. Variability in perceived duration: Pitch dynamics and vowel quality. In Carlos Gussenhoven, Yiya Chen, & Dan Dediu (Eds.) *Fourth International Symposium on Tonal Aspects of Languages*, 41–44. Grenoble: ISCA.
- Zhang, Jie. 2004. The role of contrast-specific and language-specific phonetics in contour tone distribution. In Bruce Hayes, Robert Martin Kirchner, & Donca Steriade (Eds.), *Phonetically Based Phonology*, 157–190. Cambridge: Cambridge University Press.