Tune-text negotiation: the effect of intonation on vowel duration

Martine Grice, Michelina Savino, Timo B Roettger
University of Cologne, University of Bari, Northwestern University
martine.grice@uni-koeln.de, michelina.savino@uniba.it, timo.b.roettger@gmail.com

ABSTRACT

In the variety of Italian spoken in Bari, polar questions typically have a complex rising-falling(-rising) tune, extending from the last lexical stress to the end of the phrase, whereas statements have a low-falling tune. When polar questions end in a word with final stress, the realization of the complex tune has little time to unfold and thus risks being impoverished.

We conducted a production study to investigate how speakers realize these two tunes on words with final open syllables that are stressed (sofà [so'fa] ‘sofa’) or unstressed (palo ['palo] ‘pole’). Results show that when the final vowel is stressed, its duration is considerably longer in questions than in statements (β = 111ms). We interpret this to mean that the realization of the intonation contour is facilitated by extra duration on the vowel, thus providing evidence for tonally conditioned adjustments to segment durations.

Keywords: Tune-text association, final lengthening, intonation, Italian,

1. INTRODUCTION

It is generally agreed that one of the most important properties of intonation is that it is independent of the words and phrases it occurs on [1, 19, 20]; in principle, any tune can be produced on any text. This separation of tune and text has led to important insights into the structure of intonation and the functions it expresses. Nonetheless, it has also been acknowledged that intonational and segmental aspects of speech are inherently connected. For intonation to be produced, the vocal folds have to vibrate; for intonation to be perceived, the sounds produced should ideally exhibit periodic energy with a rich harmonic structure [3, 28]. Amongst the segments that typically make up the text, some do not have vocal fold vibration at all, and others do not have enough periodic energy for intonation to be adequately perceived. The shortage of pitch-bearing material can lead to cases in which a communicatively relevant tonal contour (e.g. distinguishing a question from a statement) might be impoverished, potentially obscuring a functional contrast typically signalled by intonation.

When the requirements of a tune are not met by the text, we can talk of a tune-text conflict. Many languages have been reported to resolve such a conflict by making adjustments to the tune. Pitch targets can be undershot or fail to be realised [12, 13], referred to as partial or complete truncation, or the velocity of pitch movements can be increased, referred to as rate adjustment [8], or compression [2, 12]. These two strategies can also be combined [18, 22], and can also entail a shift in the timing of the tune, starting it earlier [18].

More recently, attention has also been paid to adjustments to the text. Pressure to realise complex tonal movements can lead to the insertion of non-lexical vowels: see, for example, [16, 23] on Tashlhiyt Berber, and [7, 9, 10] on varieties of Portuguese. In an extensive survey across a wide variety of languages, [24] point out that in contexts in which phonologically relevant tonal movements need to be realized, not only are non-lexical vowels more often inserted, but also lexical vowels are less likely to undergo devoicing and deletion.

In Italian, the language under investigation here, intonation alone bears the functional load of distinguishing between polar questions and statements, there being no morphological or syntactic differentiation. If the message is to be accurately transmitted, it is imperative that the intonation encodes this distinction: it needs to be clear whether the speaker is stating or asking something. In the variety spoken in Bari, polar questions are typically produced with a *rising* pitch accent (L+H*) on the stressed syllable, followed by a *falling* (L-L%) or *falling-rising* (L-H%) boundary tone sequence, resulting in *rise-fall* and *rise-fall-rise* contours respectively [15, 27]. Both the *rise-fall* and the *rise-fall-rise* are dynamic, complex contours, requiring a considerable amount of time to be realised.

Regardless of focus structure, the distinctive intonation contour in Bari Italian is produced on the final word in the phrase [15]. These words can have different lexical stress patterns, including antepenultimate, penultimate (the most common) and final stress. In this latter case, a complex intonation contour may be difficult to produce, especially since word-final stressed syllables in Italian are short [4, 25].

In questions ending in words with final stress, then, the challenge for speakers of Bari Italian is to indicate that they are asking a question using a distinctive intonation contour, since the required contour would need more space than a short syllable would be able to provide. [17] showed for this variety that in monosyllabic names ending in a consonant (e.g. Dick) there is a stronger tendency in polar questions to add a schwa (/dikːə/) – and a longer schwa – than otherwise, thus providing additional space to produce the complex contour over a longer stretch of pitch-bearing material.

The present paper looks at how complex contours are realised when a schwa is not likely to be introduced, i.e. when words end in a lexical vowel rather than a consonant. When the final stressed syllable is too short, some varieties employ a different tune altogether. For example, in Palermo Italian, polar questions have a *rise-fall* tune in words with penultimate stress (*domani* /doˈmaːni/ 'tomorrow') but only a *rise* in words with final stress (*falò* /faˈlɔ/ 'bonfire') or in monosyllables (*tu* /tu/ 'you') [13]. However, the variety of Italian investigated here has been reported to resolve this problem by only partially truncating the tune (*rise-slump* instead of *rise-fall* or *rise-fall-rise*), as well as by making durational adjustments to the lexical vowel [14].

The aim of the current study is to investigate the extent to which speakers make these durational adjustments, comparing the duration of word-final vowels in neutral statements, in which the intonation contour is expected to be a simple *fall*, with the same vowels in polar questions, in which the intonation is expected to be a complex *rise-fall* or *rise-fall-rise*.

2. METHOD

2.1. Participants

16 native speakers of Bari Italian (14 females, 2 males, aged 20-35) participated in the recording sessions as volunteers. They were all undergraduate and graduate students of the University of Bari. None of them had a background in phonetics or prosody.

2.2. Stimuli and Materials

Materials consisted of 32 target disyllabic words, 16 with penultimate stress (trochees) and 16 with final stress (iambs). Final vowel types were distributed evenly across the two lists, which were also balanced for word frequency and occurrence of proper and common nouns. Words were presented in two sentence modalities (question, statement), resulting in 1024 data points in total.

Contexts for eliciting target words involved mock dialogues. For the polar question context, either a positive or a negative contextualising answer was randomly selected, as in (1), whereas for the statement context the contextualising question was always an open question, as in (2). Only the target sentences were analysed (in bold).

(1) **Hai disegnato la casa?**
Have you drawn the house?
Si ho disegnato quella. / No, ho disegnato l'albergo.
Yes, I have drawn that. / No, I have drawn the hotel.

(2) Che cosa hai disegnato?
What have you drawn?
Ho disegnato la casa.
I have drawn the house.

2.3. Procedure

Speakers sat in front of a laptop screen, wearing a AKG C520 headset microphone connected to a Marantz PMD 661 digital recorder. Each target phrase was presented in context on the screen. Speakers were instructed to read each slide, first silently and then aloud at a normal rate and in a natural way. If they felt unsatisfied with their reading for any reason (either because they felt their production was unnatural, or there was a disfluency), speakers were allowed to repeat the trial, including the context. During the task, a break was inserted every 20 stimuli, but speakers were encouraged to take a break any time they felt it was necessary. Target phrases in context were presented in two separate blocks (one for each prosodic condition), and they were randomised in each block.

2.4. Annotation

Target words and final vowels were all manually segmented and annotated by a native speaker of Italian using Praat [5]. Vowel intervals were identified by discontinuities in the spectrogram along with clearly visible energy of the second formant. The end of the vowel was defined as the end of the second formant, which usually coincided with a sudden drop in the amplitude of voicing. Intonation contours were annotated holistically, with a choice of four basic contours: *fall, rise-fall, rise-fall-rise* and *rise*. Any other contours were classified as *other*.

2.5. Analysis

We fit Bayesian hierarchical regression models to measured vowel duration, using the library *brms* [6] in *R* [21]. The models include maximal random-effect structures, allowing the predictors to vary by participants and by word. We analysed vowel durations (in sec.) predicted by stress pattern (trochaic vs. iambic) and intonation pattern according to pragmatic function (*rise-fall(-rise)* in questions vs. *fall* in statements). We use weakly informative Gaussian priors (e.g. [11]), centred around 0 (sd =

0.1) for all population-level regression coefficients, as well as standard priors of the *brms* package for all other parameters. We report the means and 95% credible intervals of the estimated posteriors distributions. We consider the difference between two conditions as sufficiently compelling when the 95% credible interval of the difference does not contain 0. Data tables and R scripts can be retrieved here: http://osf.io/xhmab.

3. RESULTS

3.1. Intonation contours

The intonation contours obtained were largely as predicted from previous studies [15, 27]. Polar questions were produced with *rise-fall-rises* (L+H* L-H%) and *rise-falls* (L+H* L-L%), both in penultimate and final stress conditions. In both conditions *rise-fall-rises* prevailed, which, in this variety, is typical for the more formal, reading style [14]. See Table 1 for distributions of the different question contours across the two stress conditions.

Table 1: Distribution of intonation contours in questions across all speakers (rfr = *rise-fall-rise*; rf = *rise-fall*; r = *rise;* other).

Stress/Intonation	rfr	rf	r	other
Trochees	155	101	0	0
Iambs	135	93	25	3

It can also be seen in Table 1 that iambic words had an alternative contour some of the time. This was a simple rise, similar in shape to a continuation rise [26], and was found in the productions of three of the 16 speakers and constituted 10% of the contours on words with this stress pattern.

There were three cases of high falls (captured in the category “other” in Table 1), corresponding to H*+L L-L%, typical of contrastive narrow focus, confident checks or exclamatives. These can be regarded as negligible, given the low number and the fact that they were all produced by the same speaker.

Statements had a low fall intonation contour (H+L* L-L%) and were consistently produced across speakers and conditions. See Figures 1 and 2 for representative examples of the most common intonation contours for polar questions and statements across the two stress conditions.

Figure 1: Representative waveforms and F0 contours for polar questions (top) and statements (bottom) in trochees. Segmental annotations in SAMPA.

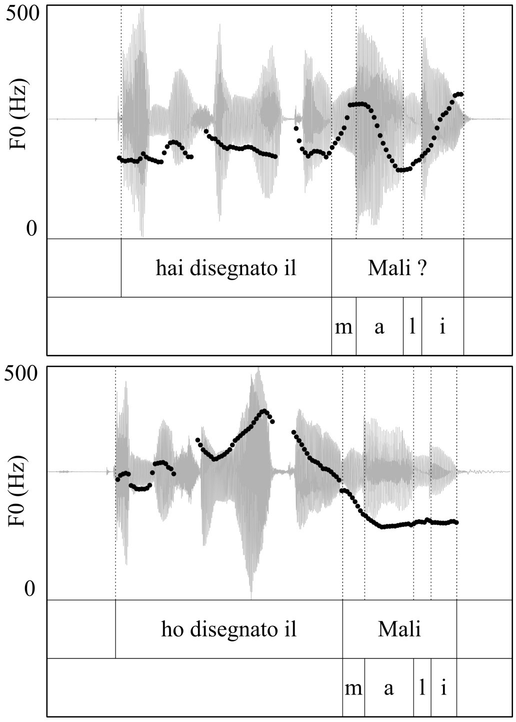
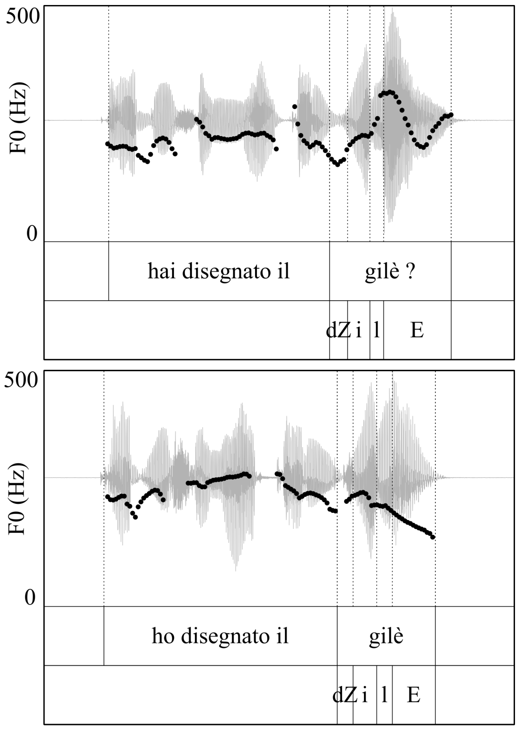


Figure 2: Representative waveforms and F0 contours for polar questions (top) and statements (bottom) in iambs.



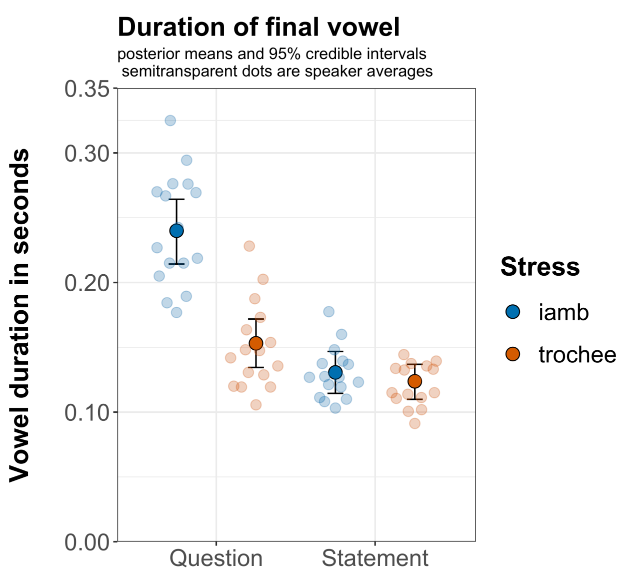
3.2. Confirmatory analyses

The vowel duration differences are shown in Table 2 and Figure 3. As predicted, there is compelling evidence for a two-way interaction between stress pattern and sentence modality in the expected direction. Trochees in questions exhibit greater durations of the vowel in the final syllable than trochees in statements (β = 29 [18,42]). The same is true for iambs but the difference is much more pronounced (β = 111 [90,130]). Generally, vowels in the final syllable are shorter when they are unstressed (trochee), but this difference is only compelling for questions (β = -89 [-110,-67]), not for statements (β = -7 [-2,7]). Even if vowels in unstressed syllables were shorter than stressed syllables in the statement condition, the difference would be very small (7ms), raising the question as to whether these differences, if credible, would be audible.

Table 2: Posterior means (*β*), 95% credible intervals, and the probability that the estimate is smaller than zero for the four relevant comparisons.

Comparison	***β***	**(95% CI)**	***Pr*(*β*<0)**
Q vs. S (trochee)	29	18,42	0.0003
Q vs. S (iamb)	111	90,130	≈ 0
Trochee vs. iamb (Q)	-89	-110,-67	≈ 1
Trochee vs. iamb (S)	-7	-21,7	0.85

Figure 3: Duration of final vowel (y-axis) as a function of stress pattern (iamb blue *left*; trochee red *right*) and sentence modality (x-axis). Semi-transparent dots indicate average values for speakers. Solid dots indicate posterior means, error bars indicate 95% credible intervals taken from the model specified above.



4. DISCUSSION

Results show that the duration of the final syllable is longer in questions than in statements. In words with final stress, the *rise-fall(-rise)* is realised on one syllable, indicating an adjustment in the duration to accommodate the complex intonation contour. This provides evidence for tonally conditioned segmental adjustments in this variety, beyond the previously attested insertion of a schwa.

We interpret these temporal adjustments to the *text* to be a result of functional pressure to produce the question *tune*. This pressure is not particularly high in trochees, as there are two syllables on which to realise the intonation contour. Nevertheless, a durational adjustment, albeit smaller, is made, indicating that the negotiation between *tune* and *text* might be sensitive to the degree of time pressure.

The use of *rise-fall-rise* or *rise-fall* does not appear to be affected by the stress pattern as both contours are used with both stress patterns, with an overall preference for rise-fall-rise, characteristic of the elicited speaking style. However, three speakers responded to the tune-text conflict by using a simple *rising* contour. A simple rise can be used to signal non-finality. This means that a clear marking of the utterance as a question may be compromised. By contrast, the majority of speakers, who always lengthen the final vowel and thereby provide a more conducive environment for the production of the complex contour, do not have to compromise conveying the intended meaning. Current investigations of non-final intonation on iambs aim to address this.

5. CONCLUSION

Our results are in line with recent production findings in a wide variety of languages concerning insertion, devoicing, and deletion of non-lexical vowels [24]. This indicates that textual adjustments for accommodating tonal needs is more common than previously predicted. If a tone or tonal complex needs to be realised, vowel insertion and lengthening can increase the segmental material that is periodic and harmonic enough to enable the articulation and perceptual retrieval of pitch movements. Likewise, reduction phenomena like vowel devoicing and deletion are less likely to occur in these contexts.

These patterns suggest that *tune* and *text* are anything but independent, and, instead, interact with each other in an intricate way to fulfil the need to transmit communicatively relevant messages. Since such a negotiation can be phonologised, these insights can broaden our understanding of how different levels of phonological representations can coevolve to uphold message transmission accuracy.

4. REFERENCES

[1] Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

[2] Bannert, R., Bredvad, A. 1975. Temporal organization of Swedish tonal accent: The effect of vowel duration. Working papers Phonetics Laboratory 10, Department of General Linguistics, Lund University.

[3] Barnes, J., Brugos, A., Veilleux, N., Shattuck-Hufnagel, S. 2014. "Segmental Influences on the Perception of Pitch Accent Scaling in English." In Proc. Speech Prosody, 7:1125–1129.

[4] Bertinetto, P., Loporcaro, M. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. JIPA, 35(2), 131-151.

[5] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

[6] Bürkner, P.-C. 2017. brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software, 80(1), 128.

[7] Cruz, M. 2013. Prosodic Variation in European Portuguese: Phrasing, Intonation and Rhythm in Central-Southern Varieties. PhD Thesis, Universidade de Lisboa.

[8] Erikson Y., Alstermark M. 1972. Fundamental Frequency correlates of the grave word accent in Swedish: the effect of vowel duration. Speech Transmission Laboratory. Quarterly Progress and Status Report 2–3, KTH, Sweden.

[9] Frota, S. 2002. Tonal Association and Target Alignment in European Portuguese Nuclear Falls. In Warner, N, Gussenhoven, C. (eds.) Papers in Laboratory Phonology 7, Berlin, New York: Mouton de Gruyter.

[10] Frota, S., Cruz, M., Castelo, J., Barros, N., Crespo-Sendra, V., Vigário, M. 2016. Tune or Text? Tune-text accommodation strategies in Portuguese. Proc. Speech Prosody 2016. Boston.

[11] Gelman, A., Jakulin, A., Pittau, M. G., Su, Y.-S., & others. (2008). A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics, 2(4), 1360-1383.

[12] Grabe, E., Post, B., Nolan, F., Farrar, K. 2000. Pitch accent realization in four varieties of British English, Journal of Phonetics, 28, 161–185.

[13] Grice, M. 1995. The intonation of interrogation in Palermo Italian: Implications for intonation theory, Tübingen: Niemeyer.

[14] Grice, M., Savino, M., Refice, M. 1997. The intonation of questions in Bari Italian: do speakers replicate their spontaneous speech when reading? PHONUS, 3, Institut fuer Phonetik/Phonologie, Univ. des Saarlandes, Saarbruecken, 1–7.

[15] Grice M., D'Imperio, M., Savino, M., Avesani, C. 2005. Strategies for intonation labelling across varieties of Italian. In: Sun-Ah Jun (ed.), Prosodic Typology: the Phonology of Intonation and Phrasing, New York: OUP, 362–389.

[16] Grice, M., Roettger, T. B., Ridouane, R. 2015. Tonal association in Tashlhiyt Berber. Evidence from polar questions and contrastive statements. Phonology, 32 (2), 241-266.

[17] Grice, M., Savino, M., Roettger T. B. 2018. Word final schwa is driven by intonation – the case of Bari Italian. JASA 143:4, 2474-2486.

[18] Hanssen, J., Peters, J., Gussenhoven, C. 2007. Phrase-final pitch accommodation effects in Dutch. Proc.XVI ICPhS, Saarbruecken, 1554-1557.

[19] Ladd, D. R. 2008. Intonational Phonology. 2nd edition [1996]. Cambridge: CUP.

[20] Pierrehumbert, J. 1980. The Phonology and Phonetics of English Intonation. Bloomington, Indiana: MIT.

[21] R Core Team. (2018). R: A language and environment for statistical computing. The R foundation for statistical computing, Vienna, URL: http://www.R-project.org.

[22] Rathcke, T. V. 2017. How Truncating Are 'Truncating Languages'? Evidence from Russian and German. Phonetica, 73: 3-4, 194-228

[23] Roettger T.B. 2017. Tonal Placement in Tashlhiyt: How an Intonation System Accommodates to Adverse Phonological Environments. Berlin: Language Science Press

[24] Roettger, T. B., Grice, M. (in press). The tune drives the text - Competing information channels of speech shape phonological systems. *Language Dynamics and Change.*

[25] Rogers, D., Darchangeli, L. (2004) Italian, JIPA, 34.1: 117-121

[26] Savino, M. 2004. Intonational cues to discourse structure in a variety of Italian. In Gilles, P. & Peters, J. (eds), Regional variation in intonation. Tübingen: Niemeyer. 161-187.

[27] Savino, M. 2012. The Intonation of Polar Questions in Italian: Where is the rise? JIPA, 42(1), 23–48.

[28] Zhang, J. 2004. "The Role of Contrast-Specific and Language-Specific Phonetics in Contour Tone Distribution." In Hayes B. Kirchner, R., Steriade, D. (eds.) Phonetically Based Phonology, 157–190. Cambridge: CUP.

ACKNOWLEDGEMENTS

We would like to thank Simona Sbranna for corpus annotation and to Mario Refice for help with scripting. Our thanks also go to SFB 1252 (German Research Foundation) for generous funding.