

Chapter 7

Describing the intonation of speech acts in Brazilian Portuguese: Methodological aspects

João Antônio de Moraes

Universidade Federal do Rio de Janeiro, CNPq

Albert Rilliard

LIMSI, CNRS, Université Paris-Saclay

This chapter details a methodology based on the notion of close-copy and equivalent-copy (’t Hart 1991), showing how systematic modifications of pitch contours using resynthesis techniques allow for testing the phonological and/or expressive nature of prosodic changes for speech-act performances. The methodology is illustrated with examples of prosodically performed speech acts in Brazilian Portuguese. A perception test shows that listeners are able to discriminate between these speech acts on prosodic cues alone. Then systematic modifications of their pitch contours are detailed and synthesized. Perceptual validation of these productions shows: (1) the relative importance of pitch versus intensity and duration for these expressions; and (2) the importance of each turning point along the stylized pitch contour, both in terms of pitch height and segmental anchoring. The results support the need for three pitch levels for the phonological description of speech acts in Brazilian Portuguese. They also show the importance of the segmental anchoring of valleys for the acceptability of contours.

1 Introduction

One of the central functions of intonation/prosody, along with phrasing and focalization, is the expression of the illocutionary force of utterances, which characterizes them as different speech acts. Classical works on the Speech Act theory



(Austin 1962; Searle 1969; Vanderveken 1990; Alston 2000) have based their inventories of speech acts mostly on the existence of performative verbs. These works have relied upon written forms of language, and account for introspective representations of hypothetical examples. Within such approaches, prosody has been only incidentally mentioned as a potential tool in speech act performances. More recent works have asserted the importance of including oral and unscripted language productions for the description of speech acts (Cresti 2005; Moneglia 2011; Raso 2012; Tamoto & Kawabata 1998). Such an approach for describing language in its actual use puts emphasis on the role of intonation in conveying speech acts. In doing so, it notably shows the relevance of prosodic variants in the performance of sets of illocutions. This is done in situations where no other factors (whether morphosyntactic or lexical) can explain the given illocutionary force that is carried by an utterance, other than the peculiarities of its prosody. In Brazilian Portuguese, for instance, a number of directive acts may be distinguished by means of their prosodic patterns; a directive act is here understood as an attempt by the speaker to induce an action from the hearer (Searle 1979).

Works such as those pursued in the framework of the C-ORAL-ROM corpus (Cresti 2005; Moneglia 2011; Raso 2012) or by other research groups (e.g., Fontaney 1991; Culpeper et al. 2003; Moraes 2012) provide significant sources of knowledge on speech act performances, especially due to the “spontaneous” nature of the data on which they base their analysis (on this notion of “spontaneous speech”, cf. Blanche-Benveniste & Bilger 1999). Because they study in detail occurrences of speech observed in context, they may derive and model their occurrence context, the relations between the speaker and the interlocutors, the speech act aims, etc. (see Kohler 2004). This allows, to some extent, for developing models of interactions based on the observation of performances unconstrained by theoretical or a priori views (but see Wagner et al. 2015, for another view on data selection). On the other hand, recognized features of such in-the-field speech productions are also characterized by low quality of the acoustic signal (compared to “lab speech”), as well as by unconstrained utterances. The use of such material makes it more difficult to perform strict comparisons of acoustic changes across studied categories. On the contrary, and as defended notably by Xu (2010), lab speech presents characteristics that are not found in unconstrained productions (such as strict control over the sources of variation, crucial for pursuing controlled psycholinguistic experimentation). Such properties of lab speech are primarily useful for (in)validating theoretical proposals, but it is desirable that various studies could be produced on the basis of typologically different types of spoken productions (Wagner et al. 2015).

We thus defend here the view that out-of-the-lab speech is of an indisputable interest in developing theories through the observation of recurrent patterns in the actual use of speech during situations of communication, and that such theoretical proposals may usefully be tested in controlled experimental settings (as has been done for example by Gussenhoven & Chen 2000; Boula de Mareüil et al. 2002; House 2005; D'Imperio et al. 2010; Vanrell et al. 2013; Borràs-Comes et al. 2015). Unconstrained recordings (in the Labovian tradition) may lead to observations that were not observed or described on the basis of constrained recordings (see for example the main perceptual correlates of smiled speech in Émond 2013, which contrast with those described by Tartter 1980). Lab speech, by its very nature of controlled recordings, reduces the variability of speech; the conclusions of studies based on such material thus have to be correlated to observations of unconstrained performances. This is notably true for inter-speaker variability (or said another way, of idiosyncratic strategies) that may be downplayed (or enhanced) by laboratory approaches based on a few subjects. Meanwhile, a careful selection of controlled parameters allows for repeatable experiments; the key here is in a number of studies that can vary some characteristics (or style, to follow the terminology in Wagner et al. 2015) of the recordings (type and structure of sentences, list and nature of speech acts, speakers' gender or dialectal origin, etc.), so as to gradually improve the understanding of the extent to which such cues may convey a given communicative function. Recurrent observations of a phenomenon back and forth in both controlled and uncontrolled circumstances are the best way to assert the robustness of theoretical observations – House (2005) being an example of such a methodology linking spontaneous and controlled utterances (see also Pešková, this volume, for similar observations).

The intonation of Portuguese has been well studied in the last decade, for both European Portuguese (EP) and Brazilian Portuguese (BP) varieties, including in its illocutionary dimensions, with works on the expressions of orders, requests, vocatives, or different kinds of questions, etc. (Moraes 2008; Frota 2014; Frota & Prieto 2015; Frota & Moraes 2016), or even from an explicit regional perspective (Frota et al. 2015). On the prosody of directive acts, and especially on the distinction between orders and requests, a number of studies have been done for BP (Bodolay 2009; Queiroz 2011; Rocha 2016; 2013), and fewer for the EP variety (Falé & Faria 2007). Previous works have shown that the prosodic patterns of such speech acts not only have different prosodic characteristics (Falé & Faria 2007), but are also recognized by hearers in a consistent way (Véliz 2004; Moraes 2008; Moraes & Rilliard 2014).

This paper describes a methodological process that allows the evaluation of the relative contribution of acoustic correlates of prosody, and the prototypicality of intonation contours, for the performance of a set of speech acts. To this aim, a resynthesis technique (implemented in Praat, Boersma & Weenink 2017), based on stylized contours, is used to systematically change some aspects of the prosodic characteristics of utterances. The perception of these controlled systematic changes allows for the discussion of modifications that challenge the interpretations of speech acts versus those that merely affect the quality of the perceived pitch contour (in the line of works such as Uldall 1960; Fónagy & Bérard 1972; and House 2005). Being decoded and interpreted by listeners, such prosodic modifications change the meaning of the verbal content, participating in the communicative meaning (Mahadin & Jaradat 2011; Nadeu & Prieto 2011; Portes et al. 2014; González-Fuente et al. 2015). Hence, by modifying the melodic contours for the pitch target's height and temporal alignment, and observing the perceptual effects of each change, we intend to show how this approach allows an evaluation of the extent to which changes are acceptable, that is, those that still carry a similar meaning as the original vs. changes that either bar access to the original speech act or change its interpretation. Changes of interpretation and acceptability address differences in the phonological meaning of prosodic contours, whereas changes of their perceived quality are merely circumstantial and/or expressive. In order to support the description of the process and give examples of possible findings, the paper bases the description of the method on a set of directive speech acts produced by speakers of Brazilian Portuguese from the Rio de Janeiro variety. As the paper focuses on the methodology and not on the linguistic description, only a restricted discussion will be given of these speech acts in this language variety (referring the interested reader to papers detailing these aspects), compared to other varieties of Portuguese, or other languages (Romance or not). Similarly, the authors believe that such an experimental process allows the researcher to tackle prosodic variation at a very early level, and then to interpret the phonological significance of some of the observed variants. For this reason, and before reaching strong conclusions, the experimental method will be applied to a variety of utterances from speakers of various regional and sociocultural origins. For the same reason, it is thought to be inconvenient to describe the observed speech acts using phonological tools before evaluating the distinctive nature of the variants.

This paper first describes a set of speech acts and their prosodic variation (Section 2.1), before presenting the concept of close-copy stylization (Section 2.2) aimed at keeping only perceptually relevant intonational variation for the tar-

geted function; other prosodic parameters are described in Section 2.3. The perceptual evaluation of systematic changes introduced in such a close-copy stylization is then presented to show how experimental processes allow for the identification of relevant patterns for the various illocutionary performances under study (Section 3). The presented data are discussed in Section 4, before concluding on the pros and cons of this approach (Section 5).

2 A corpus of seven speech acts

2.1 Phonetic description

A corpus of 462 utterances, corresponding to seven speech acts, was recorded by two Brazilian Portuguese speakers (one female and one male) from Rio de Janeiro. We chose to record two speakers so as to capture some of the inter-speaker variation linked to the varying strategies allowed in the production of such speech acts. The perceptual validation of the speakers' productions shows that all were recognized (cf. *infra*), if differences in their performances were also observed. The targeted speech acts are labelled in BP as follows: *asserção* ('statement'), *interrogação* ('yes/no question'), *ordem* ('order'), *desafio* ('threat', 'challenge'), *alerta* ('warning'), *sugestão* ('suggestion', 'advice'), and *pedido* ('request'); the Brazilian Portuguese terms will be kept in the chapter to avoid misinterpretations based on the imperfect English translations. To clarify these terms, it is worth saying, following Searle (1979), that in the speech acts *ordem* and *pedido*, speaker S wants the hearer H to perform action A. In the case of *ordem*, S has a social position of authority over H and uses it while performing this speech act. In contrast, in *pedido*, S does not use or does not have a higher social position, while performing this speech act. In the *sugestão*, S believes A will benefit H, "advising" being telling someone what is best for her or him. In the *alerta*, S believes something that may not be in H's best interest could happen in case H would not do A, so S urges H to perform A for H's own good. In the *desafio*, S doesn't want H to perform A, and dares H to do so; in this sense, it is a sort of ironic order, as S wants the reverse of what she/he says, and threatens H in case she/he performs A. Finally, in the *interrogação*, S asks for information, and in the *asserção*, S transmits this information.

The mentioned speech acts were recorded on sentences of varying length: sentences of one, two, three, six, nine, and 12 syllables were used. For each length (except the one-syllable sentence), two sentences were proposed, ending respectively with a paroxytonic or an oxytonic word. This results in a set of 11 different

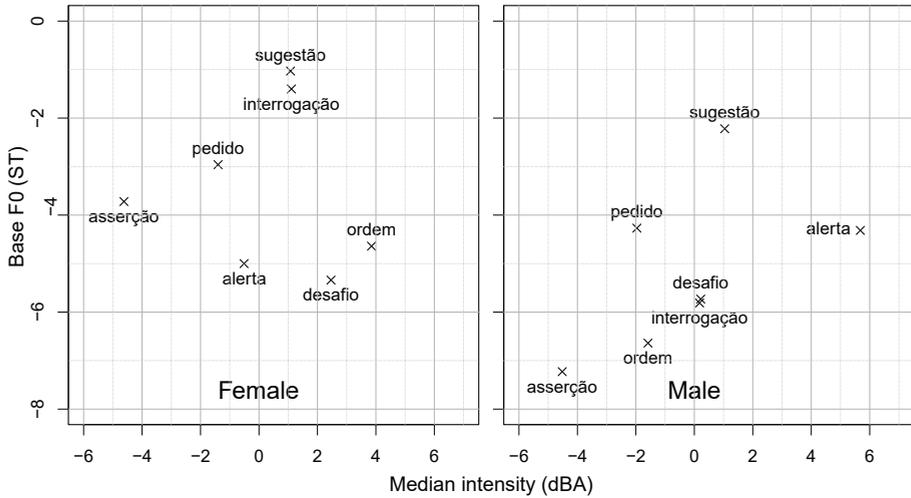


Figure 1: Position of each speech act in the (median intensity * base F0) plane, for each speaker (left: female, right: male).

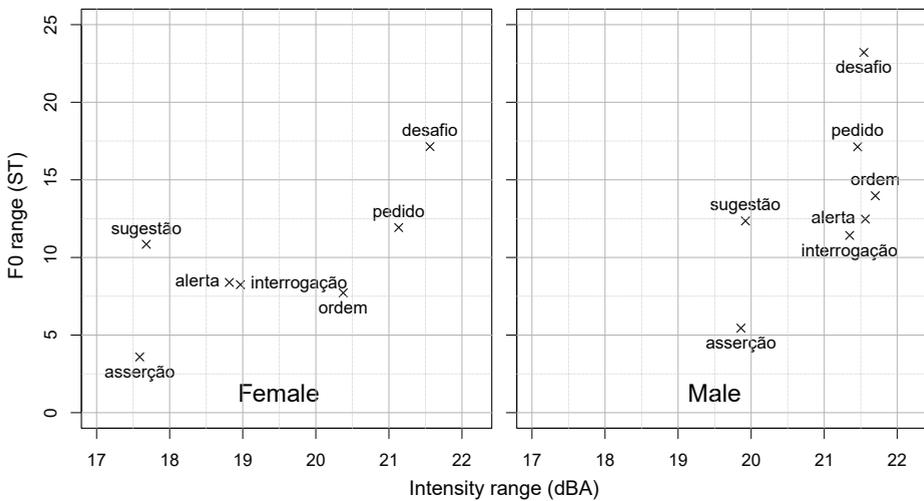


Figure 2: Position of each speech act in the (intensity range * F0 range) plane, for each speaker (left: female, right: male).

sentences (these data are based on structures similar to one that is presented in Moraes 2008, but presenting larger variation). Three repetitions of each of these sentences, for each speech act, were recorded by the two speakers. Before the recordings, each speaker was presented with the seven contexts and their communicative implications. They were then recorded, with the speaker and an experimenter monitoring the quality of the production, and recording again in case they were not satisfied with the performance.

The fundamental frequency (F0, expressed in semitones) and A-weighted intensity (expressed in dBA) of all the sentences were extracted for each 10 milliseconds of the recorded stimuli, using Praat (Boersma & Weenink 2017); values in unvoiced parts were discarded. A-weighted intensity was selected as a good correlate of the perceived vocal effort (cf. Liénard & Barras 2013; and similar measures in Traunmüller & Eriksson 2000).

Traunmüller & Eriksson (1995) introduced the notion of F0 base value, linked to the perceived register of a speaker's voice and differentiated from the size of the F0 excursion in the voice. They estimated the base value of a speaker as 1.43 standard deviation below the speaker's mean F0. Arantes & Eriksson (2014), building on this notion, have shown that the base value is a measure of preferred F0 (the frequency the speaker is more likely to use spontaneously) that stabilizes more rapidly over time compared to, for example, measures of central tendency like the mean or the median.

The F0 base value was thus estimated for each speaker and each speech act to estimate its register. The base value was measured as the 10th percentile of a speaker's F0 distribution in a given speech act – a close approximation of Traunmüller & Eriksson (1995) calculation, which also takes into account F0's skewness (cf. Arantes & Eriksson 2014: for a slightly different choice). The amplitude of pitch excursions in each speech act was estimated by calculating the difference between the 90th and the 10th percentile of F0. The median intensity was calculated to indicate the preferred effort used to perform a given speech act; the span of intensity (the difference between the 90th and the 10th percentiles of intensity) measures variation in intensity. The F0 and intensity measures were corrected for speaker-specific value by subtracting the speaker-specific values of (respectively) base F0 and median intensity from the raw measures.

Figure 1 presents the position of speech acts according to their median intensity and F0 base value. It thus represents the preferred pitch and average effort required for each speech act. Figure 2 presents the speech act's span in pitch and intensity – thus representing the acoustic variation linked to these expressions.

Figure 2 shows that both speakers are highly coherent in the prosodic changes they produce for each speech act: The distribution of F0 and intensity spans are the same – even if the male speaker uses a reduced span of intensity compared to the female speaker. In that figure, *asserção* (‘statement’) is performed with the most reduced prosodic change on both parameters, while other speech acts require more variation. Some of this variation is essentially produced as pitch changes; this is the case for *sugestão*, which shows an F0 span of around 10 semi-tones, but intensity changes are comparable to those observed in *asserção*. Speech acts of the *ordem* type are mostly performed via variation in the voice’s strength, which could also induce pitch change in the male voice. The most extreme variation among speech acts for both speakers is represented by *desafio*, which shows very large pitch and intensity spans. *Pedido*, *interrogação* and *alerta* are in intermediate positions.

Contrary to F0 and intensity spans, both speakers show different uses of preferred pitch and effort for these speech acts (see Figure 1). A first difference is observed for *asserção*: These acts show the lowest base F0 for the male speaker, while they are at about the median pitch of the female speaker. This difference in pitch usage for the two speakers may be related to their gender, and linked to sociocultural representations. Given that she has a relatively high base pitch for *asserção*, the female speaker can lower it for the expressions of *alerta*, *desafio*, and *ordem*, despite producing them with a higher median intensity. Recall that intensity and F0 are correlated (Titze & Sundberg 1992; Liénard & Di Benedetto 1999; Traunmüller & Eriksson 2000); thus, lowering F0 while speaking louder requires a strong control on vocal folds and is expressively significant. The male speaker, already producing *asserção* near his lower (comfortable) pitch, shows an increased F0 for all expressions, including *ordem*, *desafio*, and *alerta* – also produced with a higher intensity. Note that both speakers also differ in the way they perform *ordem* and *alerta*: The female’s expressions of *ordem* are louder (*ordem* also representing the loudest expression) than *alerta* while the reverse is observed for the male speaker (*alerta* being the loudest expression). One may speculate that these speech acts (*ordem*, *alerta* and *desafio*) are produced with a voice louder than *asserção* by both speakers—and that this louder voice leads mechanically to a higher pitch. This rise in pitch can be controlled by the female speaker, who uses for her neutral voice a rather high pitch, but not by the male speaker, thus limiting his pitch rise to the constraint given by a stronger effort. The increase in base F0 observed for the male speaker along these speech acts follows almost linearly the increase in intensity.

Comparatively, the increase in base F0 (from *asserção*) is more important than the increase in intensity in the cases of *sugestão* and *pedido*. *Sugestão*, which is performed with large pitch spans but few intensity changes, shows the highest F0 baseline (i.e., the highest “register”) for both speakers, with an intensity around their median value. *Pedido* is halfway between *asserção* and *sugestão*. These two expressions may thus be focused on high pitch as an expressive marker, rather than on loud voice, as could be the case with *alerta*, *ordem*, and *desafio*. The expression of *interrogação* shows an increase in base pitch and intensity over statement for both speakers, but the pitch increase is more marked in the case of the female speaker.

These average values over several sentences allow one to understand some of the differences between the seven speech acts, but they do not capture the dynamic of pitch change along the linguistic structure of sentences. In order to have a better understanding of melodic contours and to see how they may be distinctive, the next sections will focus on the description of one sentence. The verb-object, six-syllable-long (six syllables only because of the crasis between the two /a/ at the middle of the sentence) and two-accent sentence *Destranca a gaveta* (‘Unlock the drawer’), as produced by the female speaker, is used here to display the prosodic features linked to these speech acts. From the F0 estimations, close-copy stylizations of the intonation (’t Hart 1991) were hand-produced, thanks to Praat “modification” objects. Figures 3 to 9 present the close-copy obtained for one sentence, superimposed on a spectrogram, and a plot of the raw F0 values.

2.1.1 *Asserção* (‘statement’)

A sentence produced as a simple declarative *asserção* (see Figure 3), without any further expressive variation, shows a reduced F0 span (the middle 80% range spans 3.6 semitones for the female and 5.4 for the male speaker on the complete corpus), rising up to the first stressed syllable and then falling to the end of the sentence. In this configuration, the final stressed syllable bears a low level of F0 with a falling configuration, while the pre-stressed syllable has a higher pitch level. Post-stressed syllables show the lowest pitch levels, with lowest intensity and possibly devoicing or creak.

2.1.2 *Interrogação* (‘Yes/No question’)

The sentence produced as an interrogative speech act (*interrogação* – see Figure 4) shows a wider F0 span (the middle 80% range spans 8.2 semitones for the female and 11.4 for the male speaker on the complete corpus), presenting a double rising

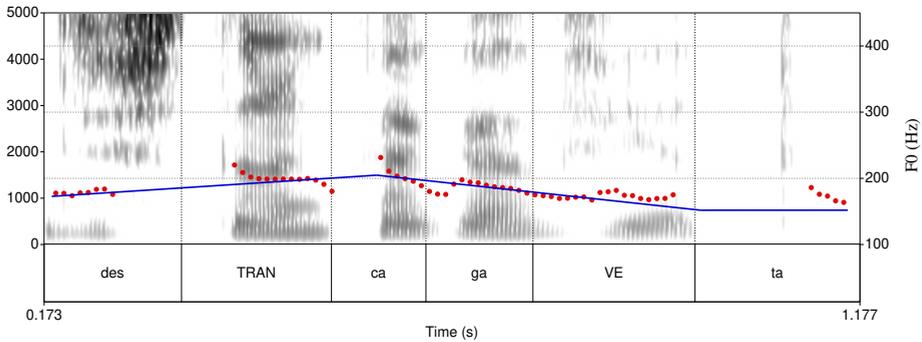


Figure 3: Representation of the sentence *Destranca a gaveta*, produced as a speech act of the *asserção* type – as in an answer to *O que ele faz quando chega em casa?* (‘What does he do when he gets home?’): Spectrogram of the recorded sound (scale in Hertz on the left), measures of F0 estimated from the signal indicated by red dots and close-copy stylization by continuous straight blue lines (in Hz, right scale); the syllabic segmentation is indicated on the bottom tier.

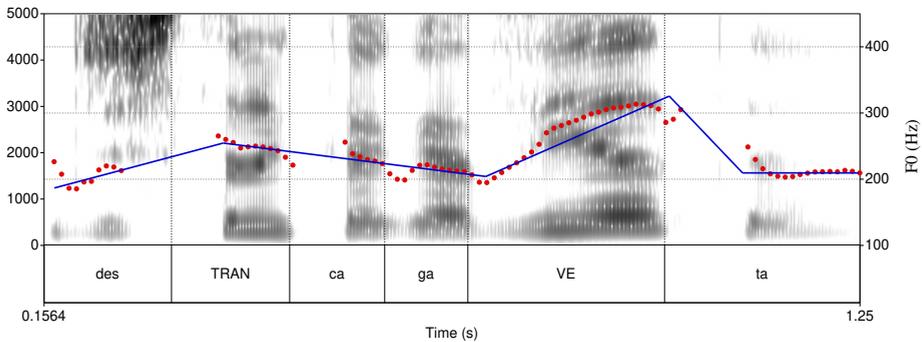


Figure 4: Representation of the sentence *Destranca a gaveta*, produced with a speech act of *interrogação*: Spectrogram of the recorded sound (scale in Hertz on the left), measures of F0 estimated from the signal indicated by red dots and close-copy stylization by continuous straight blue lines (in Hz, right scale); the syllabic segmentation is indicated on the bottom tier.

movement: The first (and the smallest) in the prenuclear position, and a second one in the nuclear position, reaching the highest level of the sentence in a rising movement along the final stressed syllable (not falling, like in *asserção*) – that is, the F0 peak has a late alignment. F0 then goes down on the potential post-stressed syllables. The valley between these two peaks is at its lowest at the end of the pre-stressed syllable.

2.1.3 *Ordem* ('order')

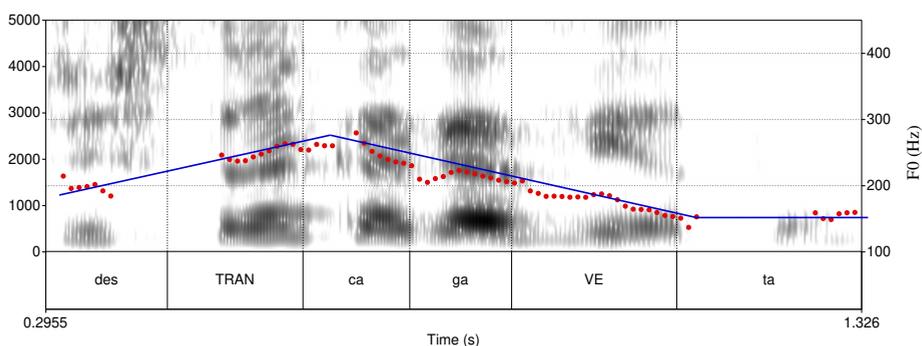


Figure 5: Representation of the sentence *Destranca a gaveta*, produced as a speech act of *ordem*: spectrogram of the recorded sound (scale in Hertz on the left), measures of F0 estimated from the signal indicated by red dots and close-copy stylization by continuous straight blue lines (in Hz, right scale); the syllabic segmentation is indicated on the bottom tier.

The expression of *ordem* (Figure 5) shows a rising–falling movement: F0 rises along the prenuclear part, reaching the sentence’s highest level on the first stressed syllable, and then falls until the nuclear position; the pitch span reaches medium levels (the middle 80% range spans 7.7 semitones for the female and 14.0 for the male speaker on the complete corpus). The final stressed syllable is performed with a low pitch and a falling configuration, while the pre-stressed syllable has a higher pitch. Potential post-stressed syllables have the lowest F0 values.

2.1.4 *Desafio* ('threat, challenge')

Desafio, as *ordem* and *asserção*, also shows a global rising–falling F0 contour (Figure 6): The movement rises during the prenuclear part, up to the end of the first stressed syllable, and goes down until the nuclear position. Differing from *as-*

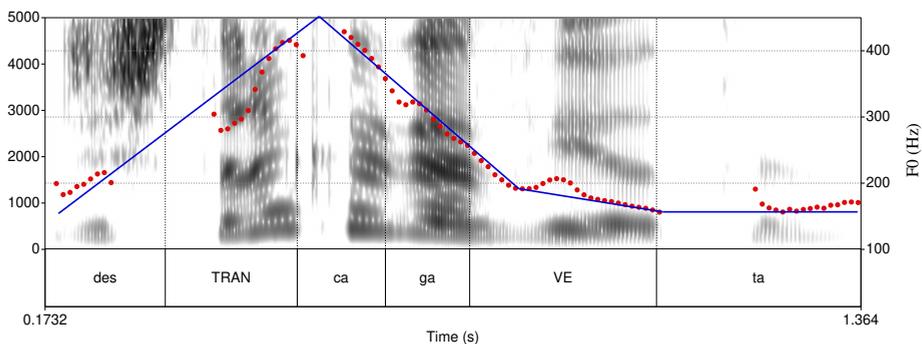


Figure 6: Representation of the sentence *Destranca a gaveta*, produced as a speech act of *desafio*: Spectrogram of the recorded sound (scale in Hertz on the left), measures of F0 estimated from the signal indicated by red dots and close-copy stylization by continuous straight blue lines (in Hz, right scale); the syllabic segmentation is indicated on the bottom tier.

serção and *ordem*, *desafio* receives the largest pitch span (the middle 80% range spans 17.1 semitones for the female and 23.2 for the male speaker on the complete corpus), with an especially high rise in the first stressed syllable. The final stressed syllable shows a falling configuration, at the end of the falling contour. The eventual post-stressed syllables receive the lowest F0 values. The change in the slopes of the falling contour on the pre-stressed syllables, compared to the stressed syllable, is not relevant as far as speech acts are concerned.

2.1.5 *Alerta* ('warning')

The intonation contour for the expression of *alerta* (Figure 7) may also be summarized as a rising–falling one, but it does not follow the same timing as those observed for *asserção*, *ordem* and *desafio*: Unlike those, the rise for *alerta* starts later, after the first stressed syllable, and ends just before the final stressed syllable. The F0 movement spans a notable F0 range (the middle 80% range spans 8.4 semitones for the female and 12.5 for the male speaker on the complete corpus). The pitch movement on the final stressed syllable is descending down to the post-stressed syllable, with a sharp fall at the beginning and an important lengthening of the stressed syllable (with a median duration in *alerta* of 0.551 and 0.604 seconds for the female and male speakers respectively, vs. median durations in *asserção* of 0.308 and 0.253 seconds for the female and male speakers). The potential post-stressed syllable continues this descending movement down to the lowest F0 in the sentence.

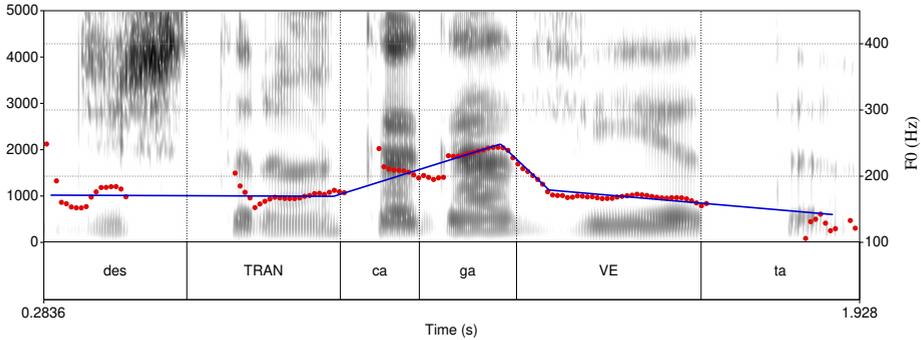


Figure 7: Representation of the sentence *Destranca a gaveta*, produced as an *alerta* speech act: Spectrogram of the recorded sound (scale in Hertz on the left), measures of F0 estimated from the signal indicated by red dots and close-copy stylization by continuous straight blue lines (in Hz, right scale); the syllabic segmentation is indicated on the bottom tier.

2.1.6 *Sugestão* ('suggestion')

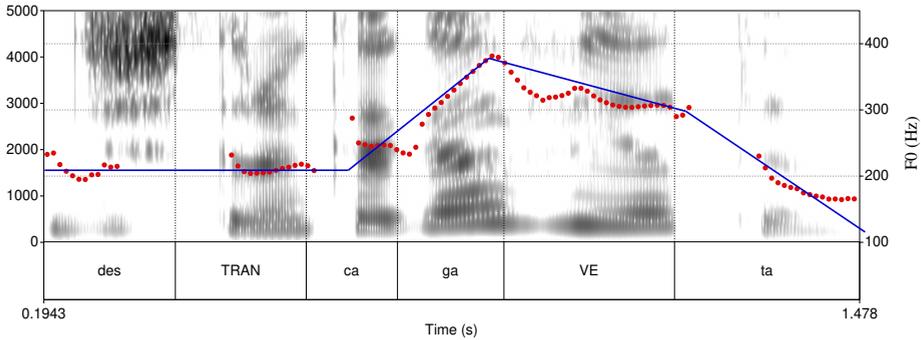


Figure 8: Representation of the sentence *Destranca a gaveta*, produced as a speech act of *sugestão*: Spectrogram of the recorded sound (scale in Hertz on the left), measures of F0 estimated from the signal indicated by red dots and close-copy stylization by continuous straight blue lines (in Hz, right scale); the syllabic segmentation is indicated on the bottom tier.

Expression of *sugestão* (Figure 8), like that for *alerta*, presents a rising–falling contour with a late rise, spanning the two unstressed syllables between the stress in prenuclear position, up to the final one. The melodic movement is wider than

for *alerta* (the middle 80% range spans 10.8 semitones for the female, and 12.4 for the male speaker on the complete corpus). After the F0 peak, the melodic movement shows a light fall on the stressed syllable that stays at a high level, and a steep fall on the post-stressed syllable, departing from *alerta*. *Sugestão* also departs from *alerta* in terms of lengthening – lacking the typical lengthening on the final stressed syllable observed in *alerta*.

2.1.7 *Pedido* ('request')

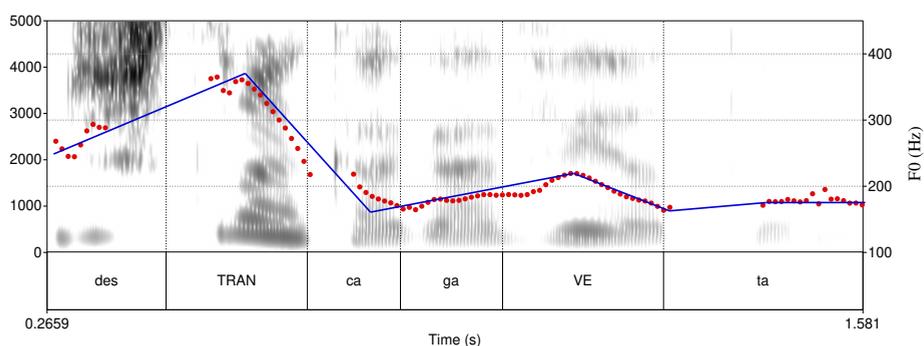


Figure 9: Representation of the sentence *Destranca a gaveta*, produced as a speech act of *pedido*: Spectrogram of the recorded sound (scale in Hertz on the left), measures of F0 estimated from the signal indicated by red dots and close-copy stylization by continuous straight blue lines (in Hz, right scale); the syllabic segmentation is indicated on the bottom tier.

Finally, the expression of *pedido* depicts a double rise (one on each stressed syllable, see Figure 9), and is similar in that respect to the *interrogação* expression. The first rise, on the stressed syllable at the prenuclear position, reaches the highest level – contrary to the *interrogação* contour. The second rise is located on the nuclear stressed vowel. Both peaks are located at the beginning of the vowels (early alignment), thus leading to a falling F0 contour during the stressed vowels. The valley between these two peaks is at its lowest at the end of the first post-stressed syllable; this also differs from the *interrogação* contour, with a much steeper slope in *pedido*, spanning the first stressed syllable and its post-stressed syllable, followed by a shallower rise up to the nuclear stress. The final post-stressed syllable bears a low tone. The F0 span is of 12.0 semitones for the female and 17.1 for the male speaker on the complete corpus (middle 80% range).

2.2 Close-copies of intonation contours

In order to explore similarities and differences between these contours, the notion of close-copy, developed by 't Hart and colleagues at IPO was used ('t Hart 1991). It basically consists of a straight-line stylization of the raw F0 contour, using as few straight lines as needed so as to obtain a perceptually indistinguishable stimulus. Close-copies were obtained here using the stylization capabilities of Praat (its "Stylize pitch" function, used with a threshold of two semitones), and then reaching a minimal number of straight lines by hand, thanks to a careful process of listening. This concept of close-copy is the basis of other processes of intonation stylization, which reach perceptually identical stimuli by means of, for example, quadratic splines (for MOMEL, cf. Hirst & Espesser 1993) or straight-line stylization of vocalic nucleus (the model of tonal perception proposed by d'Alessandro & Mertens 1995, implemented into the *Prosogram* by Mertens 2004). The lines that constitute the close-copies of each of these seven expressions are presented in Figures 3 to 9 (the straight lines that approximate the raw F0 estimation). The notion of equivalent-copy – the intonation contour of a close-copy simplified even more so as to reach a stimulus that, if not perceptually identical to the original, does not lead to any difference in its interpretation – is then used to simplify further the contours of these expressions.

The phonetic description of these seven speech acts (cf. the preceding part) leads to three types of contours:

1. rising–falling contours with a peak near the end of the prenuclear stressed syllable (and thus a rise during the first stress and a slope on the second stress);
2. rising–falling contours with a peak reached at the end of the pre-stressed syllable (nuclear stress) – and thus a rise along the unstressed part of the sentences between the stressed syllables; and
3. two-peak contours, observed on the two accented syllables (the details of these two-peak contours will be analyzed later).

These three categories of contours present equivalent-copies that share striking similarities – as well as some differences.

The first category regroups *asserção*, *ordem*, and *desafio* (Figure 10). They are distinguished mostly by the span of the F0 peaks, as one can observe in Figure 10, representing the time-normalized equivalent-copies of these three speech acts. The similarities of the three contours are obvious, as are their clear differences

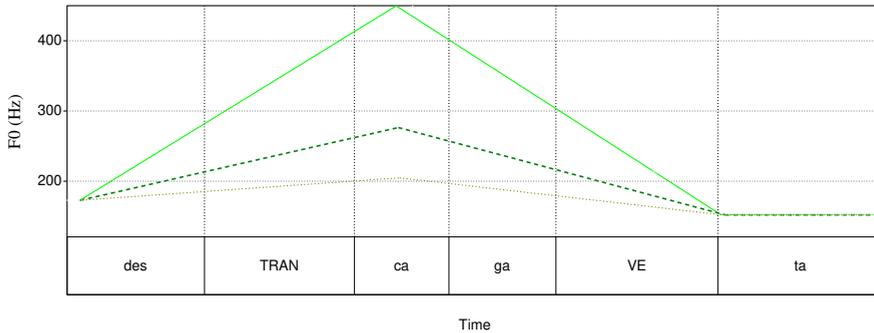


Figure 10: Equivalent-copies of the speech acts of *asserção* (dotted curve), *ordem* (dashed curve), and *desafio* (plain lines); the alignment of contours onto the phonetic segments is normalized to remove variation in timing between the expression and projected onto the segmentation of *asserção* (time-normalization is done at the level of syllable).

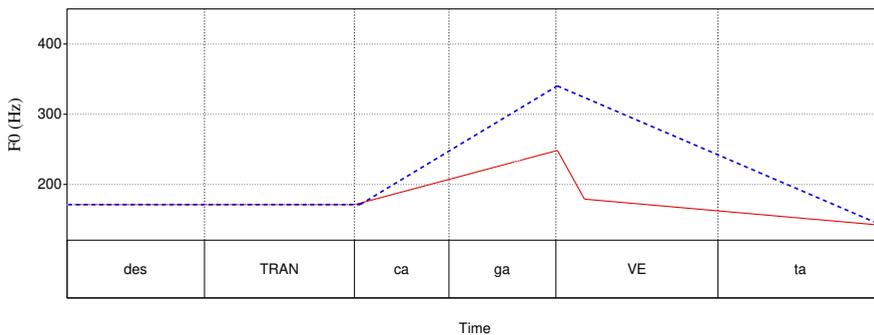


Figure 11: Equivalent-copies of the speech acts of *sugestão* (dashed curve) and *alerta* (plain lines); the alignment of contours onto the phonetic segments is normalized to remove variation in timing between the expression and projected onto the segmentation of *asserção* (time normalization is done at the level of syllable).

in terms of F0 spans. The alignment of the contours on the segmental content is comparable, as are the initial and final F0 levels of each contour.

The second category regroups the *sugestão* and *alerta* expressions, both of which have a rising–falling contour with a late rise, starting after the pre-nuclear stress, up to the nuclear stress. Figure 11 shows their two equivalent-copies. One can observe that the two rises differ, as between the expressions in the first group, in terms of pitch span but not in their segmental anchoring. The main dif-

ference, though, is linked with the falling parts, which mostly span the nuclear stressed syllable. *Sugestão* shows a continuous straight line from the beginning of the stressed syllable down to the end of the sentence, while *alerta* has a typical sharp F0 fall at the beginning of the stressed syllable (mostly on the consonant part), followed by a shallow slope through the stressed vowel, down to the end of the sentence.

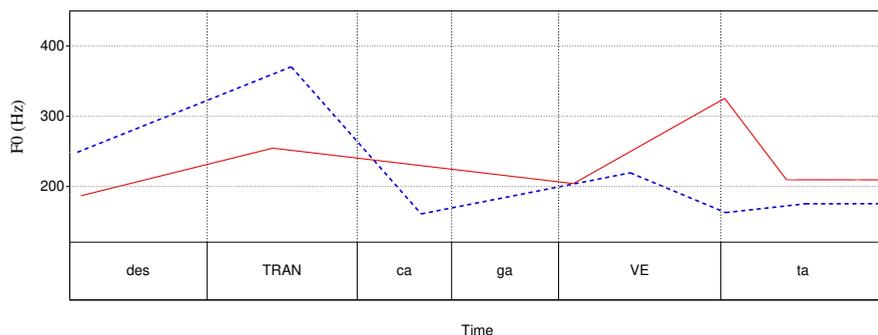


Figure 12: Equivalent-copies of the speech acts of *pedido* (dashed curve) and *interrogação* (plain lines); the alignment of contours onto the phonetic segments is normalized to remove variation in timing between the expression and projected onto the segmentation of *asserção* (time normalization is done at the level of syllable).

The third category, with *interrogação* and *pedido*, regroups two-peak contours, one on each stressed syllable (see Figure 12). Apart from the presence of peaks on these two syllables, the contours have many differences, including differences in pitch span (first peak higher for *pedido*, second for *interrogação*) and differences in segmental anchoring (the second peak of *interrogação* has a late alignment, while *pedido* is aligned to the start of the stressed vowel; the valley between the two peaks reaches its minimum at the start of the nuclear stressed syllable for *interrogação* but during the first prenuclear post-stressed syllable for *pedido*).

The determination of the equivalent-copies underlines the potential interest of two types of parameters in these performances: (1) the number of pitch levels; and (2) the segmental anchoring of peaks and valleys. These two parameters notably change the slopes, direction, and steepness on the stressed vowels.

2.3 Importance of other acoustic parameters of prosody

Equivalent-copies focus on the melodic contour as the main dimension to express such speech acts. If fundamental frequency is certainly a major parameter

for prosodic expression, intensity and duration also play a role (see Kochanski et al. 2005, for example). To that end, the use of stylization is interesting, as the stylized melodic contours can be transferred onto another sentence, with a similar segmental content but different duration, intensity, and voice quality. This is achieved by mixing, for example, a sentence’s melody – let’s say with a *pedido* pitch contour – with the *asserção* sentence’s segments, following the process schematized in Figure 13; the equivalent copy (thus the melodic contour) of *pedido* (lower left graph) is transferred onto the *asserção* (upper left graph), keeping a proportional segmental anchoring of straight lines’ beginning and end at the syllable level. This is achieved thanks to Praat’s “Manipulation” object, and the result is presented in the right graph. The resulting sound has the temporal, intensity, and articulation patterns of *asserção*, but the melodic contour of *pedido* (note that one may also have mixed the duration and intensity parameters using a similar procedure).

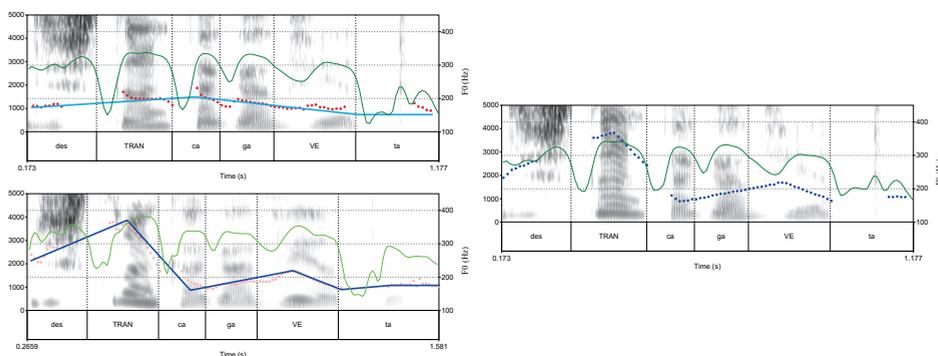


Figure 13: Schematic representation of transplanting the melodic contour of one sentence onto another: On the left, the two original sentences (top: *asserção*, bottom: *pedido*) showing the same segments but different prosodic parameters (each graph presents, over the spectrogram, the raw pitch in speckles, the intensity in curved lines, and the close-copy stylization of intonation in straight lines); on the right, the result of transplanting the melodic curve of *pedido* onto the segmental content of *asserção*.

3 Perception tests

To validate the importance of various aspects of the melodic contours described above, perception tests were carried out. The first test aimed to validate that each original performance expresses a speech act that is distinguishable from the oth-

ers and recognized for what it was intended to express; a categorical perception test was thus run on the natural stimuli, the subjects having to judge the intended speech act among the five categories (*ordem*, *pedido*, *alerta*, *sugestão*, and *desafio* – the neutral *asserção* and *interrogação* were not part of the test).

3.1 Categorical recognition

The categorical recognition test was taken by 34 subjects, native speakers of BP, who had to identify the intended speech act among the five possible acts. The stimuli were based on five sentences with varying length (one, three, six, nine and 12 syllables), each sentence produced by one female speaker with the five speech acts (thus resulting in 25 stimuli). All sentences but the one with a single syllable were based on a final word bearing a paroxytonic stress. The percentages of correct recognition (presented in Figure 14) were then analyzed (complete results, based on more speakers and expressions, are presented in Moraes & Rilliard 2014). As recognition scores are binary data, a logistic regression was applied (following Baayen 2008: 196, and using R's `glm()` procedure; R Core Team 2016). The predictors used to fit the results were the presented speech act (five levels) and the support sentence (five levels), plus the interaction between both factors.

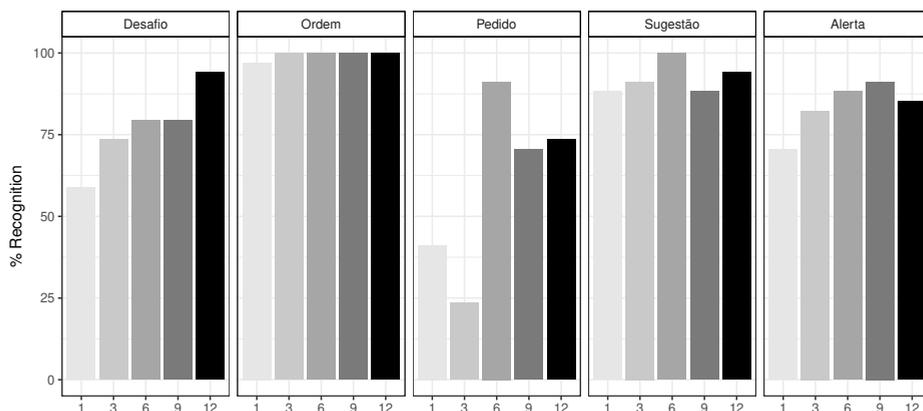


Figure 14: Bar graph representing the recognition percentages obtained by each of the 25 stimuli (five sentences performed with five speech acts); each box presents the results for one speech act, with the five sentences ranked by length.

Results of the logistic regression (see Table 1) show that all factors have a significant effect on the recognition of speech acts. The factor speech act has the largest effect, with *pedido* and *desafio* receiving the lowest recognition scores.

Table 1: Output of the logistic regression on the categorical perception results: Likelihood ratio test against the chi square distribution ($LR\chi^2$), degrees of freedom (df) and observed probability (p).

	$LR\chi^2$	df	p
Speech act	127.2	4	<0.0001
Sentence	46.8	4	<0.0001
Speech act * Sentence	27.7	16	<0.05

As one can observe in Figure 14, this is mainly due to the short sentences (one and three syllables long), and similarly, the effect of sentence length is mostly due to poor recognition scores in short sentences, mainly observed in the expressions of *pedido* and *desafio* (thus the significant interaction). From this set of sentences, the one that is analyzed in the preceding section is the six-syllable sentence, which receives high recognition scores for all speech acts.

3.2 Quality of the performance

The close-copy stylizations of the intonation pattern typical of each expression were produced from the six-syllable sentences (see Section 2). From such stylization, one may test two aspects of prosody: The perceptual importance of a given stylization parameter (in terms of melodic level or segmental anchoring) for the understanding of a given speech act, and the relative role of pitch versus intensity and duration in conveying a given expression. To test these two aspects, we will describe here a perception test that focuses on the third group of intonation contours described in Section 2.2: The two-peak contours observed for the expressions of *interrogação* and *pedido*. Several questions may be raised about these contours to have a better understanding of the parameters important for their expression:

- Minimum number of pitch levels useful for reproducing these expressions;
- Temporal anchoring of intonation contours regarding the segmental structure of the sentence;
- Importance of pitch versus the other prosodic dimensions (loudness, duration).

To test these three points, several resyntheses of the two close-copies were produced by altering various components of their structure:

- Transplanting the stylized intonation contour onto the neutral *asserção* sentence to remove any intensity and duration changes that could help in characterizing these two expressions.
- Averaging the pitch levels of the two expressions, for various parts of the contours, to test the extent to which the pitch level plays a role at a given position in the sentence.
- Changing the segmental anchoring of the pitch contours to observe whether the pitch alignment changes the understanding of expressions.

With such aims and modifications, 48 stimuli were produced (24 for each of the two speech acts), on the basis of the 12 modifications described hereafter, and resynthesized using either the original segmental information (from *pedido* or from *interrogação*), or the segmental information of *asserção*. The close-copy stylizations of *interrogação* ('yes/no question') and *pedido* ('request') are reproduced in Figure 15, with labels associated with each end of the stylized pitch segments; these points are modified in the following way:

- M1:** The original close-copy reproduction of both expressions.
- M2:** P1 averaged for frequency in both contours.
- M3:** P2 (initial peaks) averaged for frequency in both contours.
- M4:** P3 averaged for frequency in both contours.
- M5:** P4 (final peaks) averaged for frequency in both contours.
- M6:** P5 and P6 averaged for frequency in both contours.
- M7:** P3's segmental anchoring temporally averaged in both contours.
- M8:** P3's segmental anchoring inverted between the two contours (i.e., segmental anchoring of the *interrogação* sentence P3 replaced by that of *pedido*, and vice versa).
- M9:** P4's segmental anchoring inverted between the two contours (i.e., segmental anchoring of the *interrogação* sentence's final peak replaced by that of *pedido*, and vice versa).
- M10:** Averaging of frequency between both contours for each point in unaccented syllables (P1, P3, P5, P6).

- M11:** Averaging of frequency between both contours for each point (P1 to P6). Point P4b in the contour of *pedido* is given P5's averaged frequency.
- M12:** The same averaging process for frequencies as in the preceding modification, plus averaging of the segmental anchoring of unaccented syllables (P1, P3, P5, P6). The point P4b of *pedido* is removed. The two contours are similar for all their characteristics except for the segmental anchoring of peaks (P2, P4).

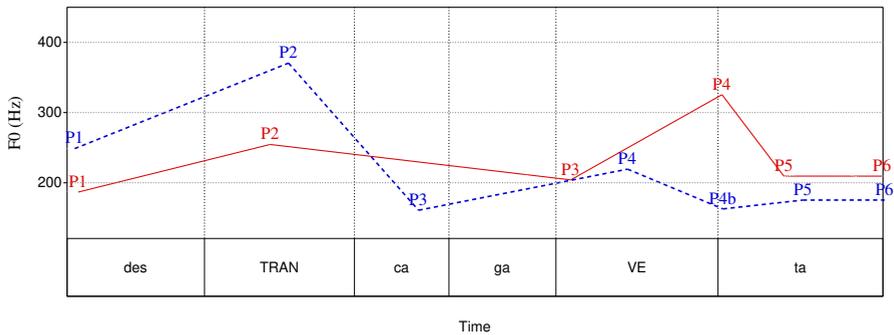


Figure 15: Target points of the close-copy stylization of *interrogação* (continuous line) and *pedido* (dashed line).

Table 2: Results of the analysis of variance run on the quality measure: Effect of each factor and interactions among factors. Columns present the degrees of freedom (*df*) of factors, the associated *F* value, the *p*-value, and effect size (η^2).

	<i>df</i>	<i>F</i> value	<i>p</i>	η^2
Modif	11	15.95	<0.0001	0.142
Speech act	1	17.95	<0.0001	0.017
Segment	1	192.22	<0.0001	0.154
Modif * Speech act	11	3.82	<0.0001	0.038
Modif * Segment	11	0.76	0.680	0.008
Speech act * Segment	1	2.16	0.142	0.002
Modif * Speech act * Segment	11	2.83	<0.01	0.029
Residuals	1056			

The 12 modifications (on two segmental substrates) obtained for *interrogação* (and *pedido*) were then presented to 23 subjects – speakers of Rio de Janeiro BP—who had to judge if each stimulus was a good performance for *interrogação* (and *pedido*), on a scale of 1 to 5. An analysis of variance was then run on these scores, with three factors: The type of modification imposed on the stimulus (12 levels), the segmental substrate used to synthesize the stimulus (two levels) and the original speech act (two levels). Interactions (double and triple) between these factors were also tested. The results are presented in Table 2.

The mean values of the quality judgments given to each stimulus are depicted in Figure 16. The reference level of a good quality for each expression is given by the two close-copies (M1) on the original segmental substrate. Post-hoc Tukey tests allowed for testing the significance of differences in quality perceived between pairs of modifications; they show the following significant differences:

- Among stimuli based on the original segmental substrate:
 - For *interrogação*: M8, M11, and M12 modifications received significantly lower scores than M1.
 - For *pedido*: M8, M9, M11, and M12 modifications received significantly lower scores than M1.
- Among stimuli based on the neutral substrate:
 - For *interrogação*: M5, M8, M11, and M12 modifications received significantly lower scores than M1 neutral.
 - For *pedido*: No modification received a significantly different score compared to M1 neutral.
- Between original and neutral-based M1 stimuli:
 - The difference between the two *interrogação* M1 modifications is not significant.
 - The difference between the two *pedido* M1 modifications is significant.

4 Discussion

Section 2 describes three categories of pitch contours, for speech acts in the Rio de Janeiro variety of Brazilian Portuguese. The first category bears a peak at the end of the prenuclear stress. Among the three performances of this category, *asserção* conveys a simple statement without imposition of the speaker onto the

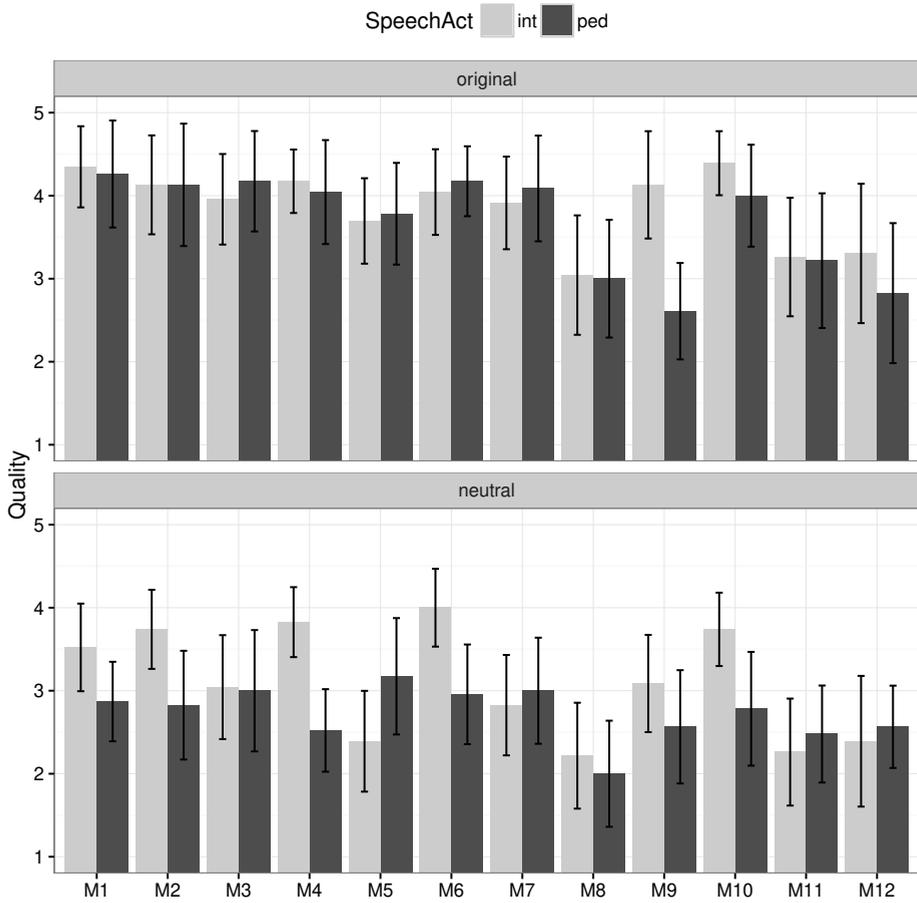


Figure 16: Bar graphs representing the mean quality given to each of the 48 stimuli (two speech acts with 12 modifications, resynthesized on twop segmental bases).

hearer. By expressing *ordem*, the speaker imposes on the listener her/his will. In *desafio*, the speaker dares the listener to perform something; the speaker does not want the hearer to perform this act. This could be interpreted under the terms of Brown & Levinson (1987) as three speech acts with varying threats to each interlocutor's face: No threat for *asserção*; a threat to the negative face of the hearer with *ordem*; and, with *desafio*, a threat by the speaker to the negative face of the hearer to dare, challenging the positive face of the speaker. For the speech acts of *ordem* and *desafio*, the speaker and the hearer are in a relation of dominance, with the speaker having power over the hearer (Spencer-Oatey 1996). The second category regroups two contours with a peak observed later in the sentence (reaching its maximum at the beginning of the nuclear stress). These contours express directive speech acts (*sugestão* and *alerta*) with a speaker in a retracted position, showing less personal implication in the outcome compared to the preceding speech acts (there is no threat on the speaker's face in these cases). The third category regroups contours performed with two peaks, one at the prenuclear and one at the nuclear position. These contours express two speech acts (*interrogação* and *pedido*) in which the speaker is somehow dependent on the listener. This could be interpreted as the speaker being in a lower position or having less power than the hearer (Spencer-Oatey 1996). For *interrogação*, the speaker is at least dependent on the hearer for having information (see the interpretation of a "desire for the goodwill of the receiver" given in Ohala 1994: 343).

Speech acts inside each of these three groups thus share similarities in shape and in meaning. In the first group of speech acts, there is an increased imposition of the speaker onto the hearer, linked to a more dominant behavior or position. This seems to be reflected in the pitch span – the wider the span, the stronger the imposition, following the mechanism of sound symbolism described by the *Effort Code* for authority (Gussenhoven 2004). The two speech acts of *ordem* and *desafio* are also performed with a relatively low base pitch (as compared to other speech acts; see Figure 1), in accordance with the *Frequency Code*'s predictions for dominant expressions (Ohala 1994). Such a need to keep the pitch low (in comparison with "neutral" pitch) may explain the difference between the strategies of the female and the male speakers regarding *ordem*. The female speaker has a base pitch for *asserção* (i.e., the more neutral speech act) close to the mean of her base pitches, while the male speaker's base pitch for *asserção* is low among his base pitches (see Figure 1). The female can lower her pitch while producing loud *ordem*; the male speaker has to refrain from getting very loud in performing *ordem* if he wants to keep the base pitch relatively low – thus, the difference in median intensity and intensity span observed for *ordem* between these two speakers.

The importance of the localization of the maximum pitch along the sentence (early, during the prenuclear part, or late, during the nuclear part) opposes speech acts that impose more (for early peaks) or less (for late maximum peaks) on the hearer. Early peaks are typical of the involved directives of the first group and of *pedido* (the first peak of *pedido* is higher than the second, contrary to *interrogação*). Late peaks are typical of the second group (distanced directives) and of *interrogação*. Such an analysis is in conformity with the *Frequency Code*, which predicts that raised or rising pitch on an utterance would tend to carry a more submissive meaning (Ohala 1994). Distinctions inside these groups are made either in terms of power, imposition, or dominance (with the gradation of imposition levels presented in the first group) or in terms of shape for the second group (a sharp pitch fall after the peak vs. a plateau).

Then the importance of pitch alignment with the segmental substrate, the number of distinctive pitch levels, and the relative importance of other prosodic parameters (duration, intensity) required to adequately perform a given speech act have been investigated.

The relative importance of prosodic parameters in conveying such speech acts in Rio de Janeiro's variety of BP is tested in transferring the melodic contours onto the neutral segmental substrate. The scores given by listeners to the resulting stimuli show that pitch alone is sufficient to express the speech act of *interrogação* – while the duration and intensity play a significant role for *pedido*, in combination with F0. Duration is also mandatory in the expression of *alerta* (see Section 2.1.5), in which the important lengthening of the final stressed syllable goes along with the melodic contour (note that the importance of duration in this case is not formally tested in this paper). The use of other parameters for expressive speech acts, and typically intensity, is linked to the notion of *involvement* in speech described by Daneš (1994), which refers to the illocutionary strength of the associated speech acts and thus to vocal effort (Gussenhoven 2004). Intensity is known to be a primary correlate of increased vocal effort during speech production (Liénard & Di Benedetto 1999; Traunmüller & Eriksson 2000). But such changes are mostly linked to changes in the speaker's affective arousal (Goudbeek & Scherer 2010) and induce changes in the mean levels of parameters – that is, gradient differences – derived from symbolic melodic gestures rather than conventional, arbitrary (hence phonological) signs (cf. Bolinger 1986).

If not the only expressive mean of prosody, melodic contours certainly bear a great deal of the prosodic speech acts' semantics in BP (Moraes & Rilliard 2014). In doing so, F0 variation expresses an additional meaning (a prosodic meaning because it is not necessarily expressed by morphosyntactic means) that changes the

original utterance's literal meaning. This allows the speaker to address a question, a statement, an order, and so on, without resorting to lexicon. The codification of these prosodic speech acts obeys phonological rules depending on languages – if possibly derived from symbolic codes (Ohala 1983; Bolinger 1986). The validity of the results presented here is, of course, limited to the Rio de Janeiro variety (one may even call into question possible sociocultural variation). Meanwhile, this paper focuses on the methodological aspect, which may be used in further experiments to test the perception of these speech acts across dialectal varieties of BP – or EP. We may speculate, given existing descriptions (cf. the introduction) and the level of diffusion through media of Rio de Janeiro's variety of BP, that such speech acts will be recognized adequately within Brazil, showing variation comparable to the one observed within the two speakers described here (cross-gender variation, or for example variation due to personality; see Rilliard et al. 2016). Conversely, the EP varieties may show much different strategies, and possibly cross-cultural differences, in such performances – but this has to be tested.

The two speech acts analyzed in detail in Section 3.2 share a global similarity, each presenting two peaks, one on each stressed syllable. But these peaks have distinct features in terms of pitch height and segmental anchoring that allow listeners to recognize each speech act. The modifications of these characteristics introduced via a resynthesis technique allowed us to test the perceptual relevance of these features for the expression of the two speech acts (hence for this speaker).

For the speech act of *interrogação*, only modifications M8, M11, and M12 significantly lowered the perception score. For *pedido*, modification M9 also received low scores. Modification M8, by changing the segmental anchoring of the valley between the two peaks, basically smoothed the final rise of the *interrogação*'s melodic contour, and the slope of the first peak for the *pedido* contour. This shows that the F0 slope, and not only the peak anchoring, has a pertinent effect on perception. M11, averaging the pitch height of each point, removed the relative difference in height between the first and second peak. M12 removed even further information from M11, causing the two contours to be almost comparable except for the peaks' segmental anchoring. M9 inverted the segmental anchoring of the second peak; thus, the peak of *pedido* arose at the end of the stressed vowel while the peak of *interrogação* arose at the beginning of the stressed vowel. The M9 modification inverted the slope of the pitch contours on the last stressed vowel: It rose for *pedido*, and went down for *interrogação*. For the speech act of *interrogação*, the height of the final peak (compared to the first peak, the second

being higher) and the steepness of this final peak's rise appear to be critical. For *pedido*, the critical characteristics are the relative height of the first peak (compared to the final peak, the second being lower), the steepness of the first peak's slope down the initial stress's vowel, and the direction of the slope on the final stress's vowel.

These results confirm the distinctive nature of the two contours beyond their two-peak categorization. Speech acts of *interrogação* in BP are characteristically recognized by their steep rise on the final stress syllable – at least in the Rio de Janeiro variety. Expressions of *pedido*, on the other hand, are characterized by two peaks on stressed syllables (the first higher than the second), with these vowels bearing falling melodic contours. This descending slope is typical of *assertão* and strong directive speech acts, while the two-peak configuration refers to *interrogação*. This mixture of characteristics may lead to the interpretation of *pedido*. The results of the experiment also shed light on the importance of the F0 slope's steepness, a factor that is linked to the anchoring of *both* peaks and valleys. The M11 modification having a perceptual effect (linked to the relevance of pitch differences between the two peaks' height), this result supports the need for at least three tonal levels in describing such melodic contour (one for valley, and two for peaks). Before concluding on the phonological differences existing between these potential three levels, more studies must be pursued, but the presented methodology allows for such strictly controlled perception tests.

5 Conclusions

The presented methodology allows one to investigate phonological differences conveyed by prosody at the speech act level. An application of this methodology was made on a corpus of seven prosodic speech acts in the Rio de Janeiro BP variety and shows that they can be perceptually distinguished. These seven speech acts are performed with three types of melodic contours: With a peak on the first stress, with a peak before the final stress, or with two peaks. The detailed characteristics of the contours in each group (their pitch span and height, their segmental anchoring, or the shape of their melodic contours) allowed listeners to distinguish each speech act. On the basis of a methodology using close-copy stylization and systematic changes in the constituents of these stylized contours, it has then been shown that the segmental anchoring of peaks and valleys, as well as the relative levels of peaks, has critical importance for the intonation of speech acts.

Typically, it is mandatory to have three levels of pitch register to describe the melodic contours conveying varying directive speech acts. Segmental anchoring of the stylized contours' valleys, on the inter-stress syllables, has an influence on the pitch slope and is thus important for the quality of the output. The melodic contour along the stressed syllables must be falling for *pedido*, while its direction is not mandatorily rising for BP interrogations; one may find falling contours on final stress for BP questions, as in confirmative yes-no questions or wh-questions (Moraes 2008). Such results may have implications (outside the focus of this paper) for phonological models of prosody, and typically within an autosegmental-metric approach (Ladd 2008), where it is customary to stick with two tone levels (see Face & Prieto 2007; D'Imperio et al. 2010).

References

- 't Hart, Johan. 1991. F0 stylization in speech: straight lines versus parabolas. *The Journal of the Acoustical Society of America* 90(6). 3368–3372.
- Alston, William. 2000. *Illocutionary acts and sentence meaning*. Ithaca: Cornell University Press.
- Arantes, Pablo & Anders Eriksson. 2014. Temporal stability of long-term measures of fundamental frequency. In *Proceedings of the 7th International Conference on Speech Prosody*, 1149–1152. Dublin, Ireland.
- Austin, John L. 1962. *How to Do Things with Words: The William James Lectures delivered at Harvard University in 1955*. Oxford: Oxford University Press.
- Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Blanche-Benveniste, Claire & Mireille Bilger. 1999. Français parlé-oral spontané. Quelques réflexions. *Revue française de linguistique appliquée* 4(2). 21–30.
- Bodolay, Adriana. 2009. *Pragmática da entonação: a relação prosódia/contexto em atos diretivos no Português*. Tese de Doutorado em Estudos Linguísticos, Universidade Federal de Minas Gerais dissertation.
- Boersma, Paul & David Weenink. 2017. *Praat: Doing phonetics by computer [Computer program]*. Version 6.0.30. <http://www.praat.org/>.
- Bolinger, Dwight. 1986. *Intonation and its parts: Melody in spoken English*. London: Edward Arnold.
- Borràs-Comes, Joan, Rafèu Sichel-Bazin & Pilar Prieto. 2015. Vocative intonation preferences are sensitive to politeness factors. *Language and Speech* 58(1). 68–83.

- Boula de Mareüil, Philippe, Philippe Célérier & Jacques Toen. 2002. Generation of emotions by a morphing technique in English, French and Spanish. In Bernard Bel & Isabelle Marlien (eds.), *Proceedings of the 1st International Conference on Speech Prosody*. Aix-en-Provence: Laboratoire Parole et Langage.
- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage* (Studies in Interactional Sociolinguistics). Cambridge: Cambridge University Press.
- Cresti, Emanuela. 2005. Per una nuova classificazione dell'illocuzione a partire da un corpus di parlato (LABLITA). In Elisabeth Burr (ed.), *Tradizione e innovazione. Atti del VI Convegno Internazionale della SILFI*, 233–246. Firenze: Franco Cesati editore.
- Culpeper, Jonathan, Derek Bousfield & Anne Wichmann. 2003. Impoliteness revisited: with special reference to dynamic and prosodic aspects. *Journal of pragmatics* 35(10-11). 1545–1579.
- d'Alessandro, Christophe & Piet Mertens. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech & Language* 9(3). 257–288.
- D'Imperio, Mariapaola, Barbara Gili Fivela & Oliver Niebuhr. 2010. Alignment perception of high intonational plateaux in Italian and German. In Mark Hasegawa-Johnson et al. (ed.), *Proceedings of the 5th International Conference on Speech Prosody*, 1–4. Chicago, IL.
- Daneš, František. 1994. Involvement with language and in language. *Journal of pragmatics* 22(3-4). 251–264.
- Émond, Caroline. 2013. *Les corrélats prosodiques et fonctionnels de la parole perçue souriante en français québécois spontané*. Thèse de doctorat en Linguistique, Université du Québec à Montréal dissertation.
- Face, Timothy L. & Pilar Prieto. 2007. Rising Accents in Castilian Spanish: A Revision of Sp_ToBI. *Journal of Portuguese Linguistics* 6(1). 117–146.
- Falé, Isabel & Isabel Hub Faria. 2007. Imperatives, orders and requests in European portuguese intonation. In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany*, 1041–1044.
- Fónagy, Ivan & Eva Bérard. 1972. «Il est huit heures»: contribution à l'analyse sémantique de la vive voix. *Phonetica* 26(3). 157–192.
- Fontaney, Louise. 1991. A la lumière de l'intonation. In Catherine Kerbrat - Orcichioni (ed.), *La question*, 113–161. Lyon: Presses universitaires de Lyon.
- Frota, Sónia. 2014. *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. Routledge.

- Frota, Sónia, Marisa Cruz, Flaviane Fernandes-Svartman, Gisela Collischonn, Aline Fonseca, Carolina Serra, Pedro Oliveira & Marina Vigário. 2015. International variation in Portuguese: European and Brazilian varieties. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 235–283. Oxford: Oxford University Press.
- Frota, Sónia & João Antônio Moraes. 2016. Intonation in European and Brazilian Portuguese. In Leo Wetzels & Sergio Menuzzi (eds.), *The Handbook of Portuguese Linguistics*, 141–166. New York: John Wiley & Sons.
- Frota, Sónia & Pilar Prieto. 2015. Intonation in Romance: Systemic similarities and differences. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 392–418. Oxford: Oxford University Press.
- González-Fuente, Santiago, Victoria Escandell-Vidal & Pilar Prieto. 2015. Gestural codas pave the way to the understanding of verbal irony. *Journal of Pragmatics* 90. 26–47.
- Goudbeek, Martijn & Klaus Scherer. 2010. Beyond arousal: Valence and potency control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America* 128(3). 1322–1336.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, Carlos & Aoju Chen. 2000. Universal and language-specific effects in the perception of question intonation. In *6th International Conference on Spoken Language Processing (ICSLP 2000)*, 91–94.
- Hirst, Daniel & Robert Espesser. 1993. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix* (15). 71–85.
- House, David. 2005. Phrase-final rises as a prosodic feature in wh-questions in Swedish human-machine dialogue. *Speech Communication* 46(3). 268–283.
- Kochanski, Greg, Esther Grabe, John Coleman & Burton Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America* 118(2). 1038–1054.
- Kohler, Klaus. 2004. Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions. In Gunnar Fant, Hiroya Fujisaki, J Cao & Yi Xu (eds.), *From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu Zongji's 95th Birthday*, 205–214. Beijing: Foreign Language Teaching & Research Press.
- Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press.

- Liénard, Jean-Sylvain & Claude Barras. 2013. Fine-Grain Voice Strength Estimation from Vowel Spectral Cues. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association 2013 (INTERSPEECH 2013)*, 128–132. Lyon, France: Curran Associates, Inc.
- Liénard, Jean-Sylvain & Maria-Gabriella Di Benedetto. 1999. Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America* 106(1). 411–422.
- Mahadin, Radwan & Abdulazeez Jaradat. 2011. The pragmatic function of intonation in Irbid dialect: Acoustic analysis of some speech acts. *International Journal of Humanities and Social Science* 1(9). 243–251.
- Mertens, Piet. 2004. Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des Langues* 45(2). 109–130.
- Moneglia, Massimo. 2011. Spoken corpora and pragmatics. *Revista Brasileira de Linguística Aplicada* 11(2). 479–519.
- Moraes, João Antônio. 2008. The Pitch Accents in Brazilian Portuguese: analysis by synthesis. In P. A. Barbosa, S. Madureira & C. Reis (eds.), *Proceedings of the 4th International Conference on Speech Prosody*, 389–397. Campinas, Brazil.
- Moraes, João Antônio. 2012. Illocution and Intonation. In Heliana Mello, Massimo Pettorino & Tommaso Raso (eds.), *Proceedings of the VIIth GSCP International Conference. Speech and Corpora*, 43–50. Firenze University Press.
- Moraes, João Antônio & Albert Rilliard. 2014. Illocution, attitudes and prosody: A multimodal analysis. In Tommaso Raso & Heliana Mello (eds.), *Spoken Corpora and Linguistic Studies*, 233–270. Amsterdam: John Benjamins.
- Nadeu, Marianna & Pilar Prieto. 2011. Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics* 43(3). 841–854.
- Ohala, John J. 1983. Cross-language use of pitch: an ethological view. *Phonetica* 40(1). 1–18.
- Ohala, John J. 1994. The frequency codes underlies the sound symbolic use of voice pitch. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 325–347. Cambridge: Cambridge University Press.
- Pešková, Andrea. 2018. Intonation of pronominal subjects in Porteño Spanish: Analysis of spontaneous speech. In Ingo Feldhausen, Jan Fliessbach & Maria del Mar Vanrell (eds.), *Methods in prosody: A Romance language perspective (Studies in Laboratory Phonology)*, 45–79. Berlin: Language Science Press.
- Portes, Cristel, Claire Beyssade, Amandine Michelas, Jean-Marie Marandin & Maud Champagne-Lavau. 2014. The dialogical dimension of intonational meaning: evidence from French. *Journal of Pragmatics* 74. 15–29.

- Queiroz, Horácio dos Santos. 2011. *A contribuição da prosódia e da qualidade de voz na expressão de atitudes do locutor em atos de fala diretivos*. Tese de Doutorado em Estudos Linguísticos, Universidade Federal de Minas Gerais dissertation.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Raso, Tommaso. 2012. O C-ORAL-BRASIL e a Teoria da Língua em Ato. In Tommaso Raso & Heliana Mello (eds.), *C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal*, 91–124. Belo Horizonte: Editora UFMG.
- Rilliard, Albert, Donna Erickson, João Antônio Moraes & Takaaki Shochi. 2016. On the varying reception of speakers expressivity across gender and cultures, and inference in their personalities. In Sandra Madureira (ed.), *Sonoridades – Sonorities*, 149–163. Lyon: Edição da Pontifícia Universidade Católica de São Paulo.
- Rocha, Bruno. 2013. Metodologia empírica para o estudo de ilocuções do português brasileiro. *Domínios de linguagem* 7. 109–148.
- Rocha, Bruno. 2016. *Uma metodologia empírica para a identificação e descrição de ilocuções e a sua aplicação para o estudo da ordem em PB e Italiano*. Tese de Doutorado em Estudos Linguísticos, Universidade Federal de Minas Gerais dissertation.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Searle, John R. 1979. A Taxonomy of Illocutionary Acts. In John R. Searle (ed.), *Expression and Meaning*, 1–29. Cambridge, Mass.: Cambridge University Press.
- Spencer-Oatey, Helen. 1996. Reconsidering power and distance. *Journal of Pragmatics* 26. 1–24.
- Tamoto, Masafumi & Takeshi Kawabata. 1998. A schema for illocutionary act identification with prosodic features. In *5th International Conference on Spoken Language Processing (ICSLP'98)*.
- Tartter, Vivien C. 1980. Happy talk: Perceptual and acoustic effects of smiling on speech. *Attention, Perception, & Psychophysics* 27(1). 24–27.
- Titze, Ingo R. & Johan Sundberg. 1992. Vocal intensity in speakers and singers. *The Journal of the Acoustical Society of America* 91(5). 2936–2946.
- Traunmüller, Hartmut & Anders Eriksson. 1995. The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America* 97(3). 1905–1915.

- Traunmüller, Hartmut & Anders Eriksson. 2000. Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America* 107(6). 3438–3451.
- Uldall, Elizabeth. 1960. Attitudinal meanings conveyed by intonation contours. *Language and Speech* 3(4). 223–234.
- Vanderveken, Daniel. 1990. *Meaning and Speech Acts: Principles of Language Use*. Vol. 1. Cambridge University Press.
- Vanrell, Maria del Mar, Ignasi Mascaró, Francesc Torres-Tamarit & Pilar Prieto. 2013. Intonation as an encoder of speaker certainty: Information and confirmation yes-no questions in Catalan. *Language and Speech* 56(2). 163–190.
- Véliz, Mauricio. 2004. Intonational devices used in the distinction of speech acts. *Literatura y lingüística*. 211–220.
- Wagner, Petra, Jürgen Trouvain & Frank Zimmerer. 2015. In defense of stylistic diversity in speech research. *Journal of Phonetics* 48. 1–12.
- Xu, Yi. 2010. In defense of lab speech. *Journal of Phonetics* 38. 329–336.