DE GRUYTER MOUTON
DOI 10.1515/lp-2013-0008 — Laboratory Phonology 2013; 4(1): 191–219

Francesco Cangemi and Mariapaola D'Imperio

Tempo and the perception of sentence modality

Abstract: Phonological models of intonation use abstract categories, such as pitch accents, to build a bridge between continuous modulations in F_0 contours (on the substantial side) and post-lexical meaning (on the functional side). However, recent research on Romance, Germanic, and non-Indo-European languages shows that sentence modality contrasts (i.e., question vs. statement) are often realized not only with different F_0 contours, but also through differences in individual phone duration or global speech rate. If these durational differences were also used as a cue in the perception of sentence modality contrasts, phonological categories in current models of intonation would qualify as excessively underspecified, and they should be expanded in order to include phonetic information on the temporal dimension as well. In this paper we evaluate the role of durational differences as a cue to the perception of sentence modality contrasts in the Neapolitan regional variety of Italian. Read sentences were resynthesized by switching durational and intonational patterns of questions and statements, and used in a forced-choice identification task. The results show that listeners exclusively rely on F_0, thus suggesting that, at least for this specific contrast in this specific variety, phonological representations of intonational contrasts do not need to be enriched with phonetic detail at the durational level.

Francesco Cangemi: Aix-Marseille University & Laboratoire Parole et Langage (CNRS), Aix en Provence, France. E-mail: francesco.cangemi@lpl-aix.fr
Mariapaola D'Imperio: Aix-Marseille University & Laboratoire Parole et Langage (CNRS), Aix en Provence, France; Institut Universitaire de France.

1 Introduction

1.1 Prosodic detail

Most influential models of intonation posit that the interface between phonetic substance, such as fundamental frequency, and intonational meaning is mediated by abstract phonological forms. This is the case of phonological categories such as pitch accents and boundary tones in the Autosegmental-Metrical (AM) framework (Pierrehumbert 1980; see Ladd 2008 for a review) and peaks or valleys in the Kiel Intonation Model (Kohler 1991). In order to function as contrastive elements, these categories are underspecified with respect to the richness of the phonetic signal, and imply the existence of a complex phonetic implementation module (Pierrehumbert and Beckman 1998). Tunes in AM, for example, are represented as a sequence of discrete tonal events, which are in turn defined by reducing phonetic information from the fundamental frequency contour to a minimal set of phonological events (e.g., Low and High F_0 levels and tonal association with metrical positions and/or prosodic edges). That is, phonetic information in the signal is reduced to the purpose of representing contrastive elements, and this reduction is twofold: first, by concentrating on F_0 alone (Liberman 1975), and second, by stylizing continuous F_0 modulations into a sequence of low and high tones linked by (mostly) linear interpolation (Pierrehumbert 1980). Reduction of phonetic information to exclusively contrastive features in the representation of phonological categories is consistent with abstractionist approaches to speech perception and word recognition (see, among others, Levelt 1989; Norris 1994; Stevens 2002). Exemplar-based approaches, on the other hand, assume that linguistic categories emerge through the activation of highly detailed memory traces (see, among others, Goldinger 1996; Johnson 1997; Pierrehumbert 2001). Phonetic detail excluded from current abstract phonological categories could thus be readily accommodated in an exemplar-based perspective.

Recent work on phonetic variability in the realization of pitch accent is thus starting to explore the viability of an exemplar approach to prosody (Schweitzer, Walsh, Calhoun, and Schütze 2011, and references therein). However, some phonetic detail at the prosodic level could also be accommodated into an abstractionist approach such as AM, notably by enriching the phonetic specification of phonological contrasts. Between the two types of information reduction mentioned above, Petrone and Niebuhr (in press) focus on the latter, namely the role of F_0 regions stretching between tonal events. Specifically, they show that sentence modality contrasts in German are characterized by different shapes in the F_0 contour interpolating between prenuclear and nuclear accents, and that this information is used in perception as well. In this paper, we address the first type of information reduction, focusing on whether F_0 is indeed the only phonetic dimension that needs to be included in the phonological representation of intonational contrasts, or whether temporal variability should also be incorporated.

1.2 Duration and intonational meaning

Sentence modality contrasts, such as the one between declaratives and polar or yes-no questions, can be expressed by morphological (question particles and affixes), syntactic (e.g., word order), and/or intonational means in the languages of the world (Dryer 2011). For languages which exclusively rely on intonation in order to signal questionhood, such as Italian and its regional varieties, a vast amount of experimental research has illustrated the importance of F_0 modulations in both production and perception (D’Imperio and House 1997; D’Imperio 2001; Petrone 2008). As a consequence, phonological accounts of the intonational means used in signaling questions usually include F_0 information in their abstract categories. For example, AM accounts of polar questions in Florentine and Neapolitan Italian translate phonetic proprieties of the F_0 contour (final rise and peak alignment within the nuclear accent, respectively) into phonologically contrastive elements (i.e., H% boundary tone and L*+H nuclear accent; see Grice et al. 2005). More recently, F_0 height information within the nuclear accent has been found to be an additional element used to signal questionhood (Vanrell 2011).

However, recent research has shown that pairs of declaratives and polar questions are often characterized by acoustical differences involving phonetic dimensions other than F_0, in particular at the durational level. Specifically, sentence modality contrasts appear to be conveyed by durational differences in a variety of languages, ranging from Dutch and Orkney English (van Heuven and van Zanten 2005) to Spanish (Henriksen 2012; Muñiz Cachón et al. 2012), French (Ryalls et al. 1994), and Italian varieties (Maturi 1988; Petrone 2008; De Dominicis 2010), and to non-Indo-European languages such as Manado Malay (van Heuven and van Zanten 2005) and a number of languages in the Niger-Congo family (Rialland 2007). For example, Manado Malay sentences uttered as questions have a faster speech rate, essentially due to the shortening of the last foot (van Heuven and van Zanten 2005). The picture is reversed in Asturian, where questions are globally longer than declaratives (Muñiz Cachón et al. 2012); the effect is also stronger on the last vowel. Note also that some authors have criticized an exclusively acoustic approach to the study of global durational contrasts between questions and statements. For instance, Smith (2002) suggests that in French the source of this durational difference might be an epiphenomenon of a different phrasing structure in the two modalities, hence leading to a different amount of boundary-induced lengthening.

Durational effects which appear to be contrastive have also been found for other types of contours, such as the “chanted call” of English and Bengali, which has been phonologically analyzed in terms of rhythmical contrast (through grid adjustments) by Hayes and Lahiri (1991). Also, it appears that in French what is usually known as “chanted list” is realized through an F_0 mid-plateau, which can be lengthened depending on syllable number within the word (Fagyal 1997). Among Romance languages, Portuguese also possesses a form of vocative chant (the greeting call) sharing some durational characteristics of the English tune, in that a significant lengthening of the nuclear stressed as well as the boundary syllable is observed (Frota et al. to appear). It has also been recently observed that in Neapolitan Italian, contrastive topics in Subject position are marked by an LH* pitch accent sharing the same alignment properties as prenuclear LH* accents, though implemented with increased duration within the stressed syllable (D’Imperio and Cangemi 2011). This is reminiscent of the durational differences found in production for the L*+H L-H% uncertainty contour of American English, which, though sharing the same intonational properties of the incredulity contour, has been found to be longer. Note though that, among the examples cited above, only the American English uncertainty/incredulity contrast has been perceptually tested, and indeed the durational difference found in production has proved not to be perceptually salient (as opposed to pitch range variability) in the identification of resynthesized stimuli (Hirschberg and Ward 1992).

One can argue then that systematic production differences in a phonetic dimension, such as duration, between two putative intonation contours is necessary but not sufficient evidence for the difference to be perceived, hence used by listeners. In this context, an enrichment of abstract phonological representations for intonation would only be justified if the durational differences reported above were also used as a cue in the perception of sentence modality contrasts. For this reason, in this paper we first show that a durational difference exists between Neapolitan questions and statements, though its characterization is quite complex and emerges only when local segmental duration increments are taken into account (Section 2). Moreover, the perceptual role of these durational differences is tested in order to determine whether the phonetic dimension of *duration* should be included as an additional phonological dimension (*tempo*) in the characterization of intonational contrast. In particular, two views of tempo are particularly relevant to our discussion and especially useful in the development of our hypotheses, and hence they will be presented below.

1.3 Two views of tempo

In her pioneering work on prosodic or suprasegmental features, Ilse Lehiste suggested that three axes are relevant to the study of prosody, namely quantity, tonal, and stress features. For each of these three dimensions, the study of “all inherent constraints and conditioned variations” is the first step towards its evaluation as an “independent variable” (Lehiste 1970: 3). That is, phonetic knowledge (articulatory, acoustic, and perceptual) is a necessary prerequisite for the exploration of linguistic function (on both word and sentence level) and ultimately, we might add, of phonology. To exemplify, given the tonal dimension, phonetic knowledge of phonation (articulatory), fundamental frequency (acoustics), and pitch (perception) allows the exploration of tone (word level) and intonation (sentence level). Thus, for example, intonation refers to sentence-level functions of tonal features. In her account, tempo represents for the quantity dimension what intonation represents for the tonal dimension, namely a potential sentence-level function of quantity features. Our own use of tempo is rooted in this definition.

An alternative view is exemplified by AM theory, in which, for instance, intonation "refers to the use of *suprasegmental* phonetic features to convey 'postlexical' or *sentence-level* pragmatic meanings in a *linguistically structured* way" (Ladd 2008: 4; original emphasis). Hence, on one side, just as in Lehiste's definition, the domain of intonational function is the sentence level. However, in Ladd's definition of intonation, the phonetic features which are potentially relevant to intonation are not limited to tonal features, but include all suprasegmentals, namely "features of fundamental frequency, intensity and duration" (Ladd 2008: 4). As a result, the notion of intonation according to Ladd encompasses the wider spectrum of all phonetic features relevant for Lehiste's prosodic correlates.[1] However, intonation in the AM framework is also characterized by linguistic structuring, given that intonational features "exclude 'paralinguistic' features in which continuously variable physical parameters (e.g., tempo and loudness) directly signal continuously variable states of the speaker (e.g., degree of involvement or arousal)" (Ladd 2008: 6). Within this framework, tempo is thus a synonym for speech rate.

In other words, tempo is viewed as a physical parameter by Ladd and as a (potential) phonological dimension by Lehiste; however, terminological divergences are of limited interest when compared to substantial similarities. In both cases, durational differences which are not lexical in nature are seen as mainly related to paralinguistic meaning (e.g., changes in the "mood of the speaker", see Lehiste 1970: Section 2.5.3). The literature we reviewed above, on the other hand, shows that in some languages durational patterns vary as a function of either sentence modality or pragmatic meaning. This kind of evidence could be accommodated within Lehiste's view of prosody by allowing tempo to serve linguistic functions as well.[2] Nevertheless, a link between durational patterns and intona-

[1] The role in pitch accent contrasts of phonetic information other than F_0 contour has been explicitly explored in the AM framework at various stages of its development, since Ward and Hirschberg (1988) up to Zadeh, Gussenhoven, and Bijankhan (2011).

[2] Indeed, this would restore symmetry in the kind of functions exerted by the various suprasegmental features. According to the original architecture, in fact, whereas stress features and intonation map on both linguistic and attitudinal meaning, tempo maps on attitudinal meaning alone.

tional meaning could also be accommodated in the AM framework. In this case, tempo would not cue linguistic functions in itself, while durational features might enrich phonological representations of intonation. Both perspectives, however, rely on the assumption that the different durational patterns one can find in the production of sentence modality contrasts (see Section 2) are perceptible and actually used by listeners. This issue is relevant to our broader research question on the integration of phonetic detail into abstract representations. Our research hypotheses will thus be formulated so as to address both questions (see Section 3.1).

2 Production

2.1 Durational patterns in the production of sentence modality contrasts in Neapolitan Italian

As we said above, recent cross-linguistic evidence shows the existence of systematically produced durational differences across sentence modalities. However, the general picture is far from clear, since the various studies concern typologically different languages. Moreover, possible confounding factors (such as syntactic or information structure) are not homogenously controlled in the various corpora, which are also rarely composed of recordings from more than ten speakers. On top of that, acoustic measures are carried over at different degrees of precision, ranging from the duration of an entire utterance to that of individual segments. Nonetheless, the literature seems to converge towards two main results. First, there are global durational differences across statements and questions, even if the effects can take opposite directions in the various languages: for example, questions are longer than statements in Gur and Kru languages (Rialland 2007) but shorter in Manado Malay and Dutch (van Heuven and van Zanten 2005). Second, these durational differences seem to be localized at specific portions of the utterance. However, once again the various studies present disparate results in terms of the size and the position of the relevant units: questions have longer final syllables in Canadian French (Ryalls et al. 1994), a shorter "stretch between the stressed syllable on the subject and that on the object" in Dutch (van Heuven and van Zanten 2005), and speaker-dependent effects on the duration of the initial phonological word in Neapolitan Italian (Petrone 2008).

Given the mixed results in the literature, we decided to verify through three studies on Neapolitan Italian (Cangemi and D'Imperio 2011a, 2011b, to appear) the existence of different durational patterns in production, for questions and

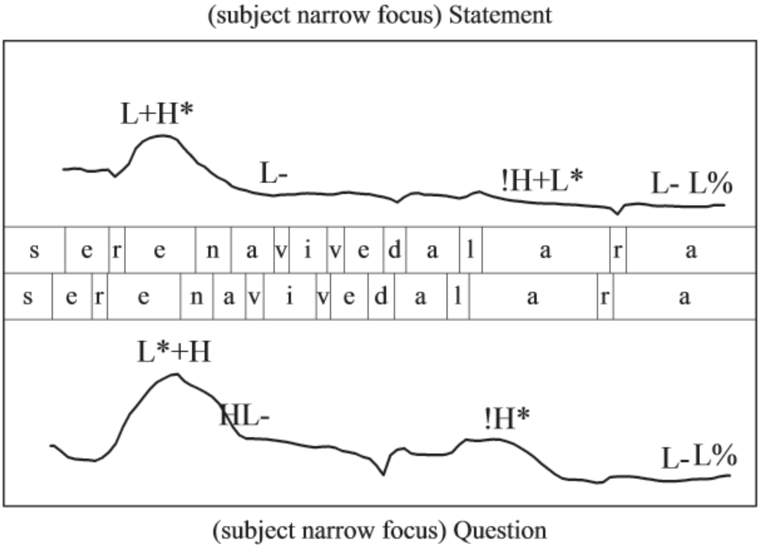

Fig. 1: F_0 contour, tonal labeling and phone segmentation for statement (top) and question (bottom) utterance of the sentence *[Serena]$_F$ vive da Lara* ‘Serena lives at Lara’s’ read by the same female Neapolitan Italian native speaker. Range 0–1 s (time, x-axis) and 150–420 Hz (F_0, y-axis).

statements characterized by the same lexical material, phrasing, and accent structure. The question/statement contrast is exemplified in Figure 1, showing F_0 contour, tonal labeling, and phone segmentation for the sentence *Serena vive da Lara* (‘Serena lives at Lara’s’, see Example 1b at Section 2.3) uttered with narrow focus on the subject and either a declarative (top panel) or a question intonation (bottom panel) by the same female speaker. Melodic differences are clear, especially within the focal pitch accent on *Serena*, whose peak is aligned closer to the stressed vowel offset in the question, as well as in the post-focal pitch accent on *Lara*, which has a steeper fall in the question.[3] Note also that the phone segmentation shows differences in the duration of the first and last segment, in that the first segment appears to be longer in the statement while the last segment is longer in the question.

[3] Following Grice et al. (2005), the first pitch accent is nuclear, and can be labeled as L+H* in the statement and L*+H in the question. The second pitch accent is postnuclear and thus takes a diacritic for range compression, and can be labeled as !H+L* in the statement and !H* in the question. Phrase accent and boundary tone at utterance right edge are L-L% in both contexts, while phrase accents after the nuclear accent are transcribed as L- for statements and HL- for questions, following D’Imperio and Cangemi (2011).

With respect to the studies available in the literature, the main improvements of this corpus can be found in the use of (1) explicitly operationalized hypotheses (see Section 2.2), along with (2) two corpora carefully balanced with respect to possible confounding factors, (3) a high number of speakers, and (4) fine-grained duration measurements (see Section 2.3).

2.2 Hypotheses

A synopsis of the findings available in the literature allowed us to formulate two general but explicit hypotheses. According to the first, *sentences have a different duration when uttered as questions or statements* (H1), independently of the direction of the effect. That is, we expect different utterance duration (U) for questions (Q) and statements (S):

H1: $U_Q \neq U_S$

The second hypothesis is that *durational differences are localized in some specific portions of the utterance* (H2), independently of their position and their size. That is, durational differences between questions and statements are not due to uniform stretching or compression of all individual segments. We expect that no single coefficient (a) could transform the duration of each phone (P) in a sentence, from first (1) to last (n), from its value in statement utterances (S) to its value in question utterances (Q).

H2: $P\{1,n\}_Q \neq aP\{1,n\}_S$

By jointly evaluating H1 and H2, we can foresee four scenarios. In the first, neither H1 nor H2 are supported by our data. In this case, there would be no evidence for different durational patterns in the production of sentence modality contrasts (null hypothesis). If only H1 is validated, we would have evidence of a global effect of sentence modality on utterance duration. This could be interpreted in terms of a global speech rate difference across modalities. If both H1 and H2 are supported, we could conclude not only that durational differences are localized in a specific portion of the utterance, but also that global duration is affected by local variations. In the fourth scenario, only H2 is supported. In this case, we would have evidence for the existence of local durational differences between questions and statements which counterbalance each other at the utterance level. This would entail the existence of at least two loci for durational differences, with opposite properties. That is, one portion of the utterance is expected to be longer in statements, and another in questions.

2.3 Method

We designed two sets of sentences. In the first (referred to as the *Orlando* set) we enforced a tight control on various possible confounding factors, while in the second (the *Danser* set) we loosened some of the constraints in order to enhance communicative plausibility. The three SVO sentences in the *Orlando* set were composed of 8 syllables, all with CV structure with voiced consonants, grouped in three words (of 3, 2, and 3 syllables) with penultimate stress. The structure was thus [CV.ˈCV.CV]$_S$ [ˈCV.CV]$_V$ [CV.ˈCV.CV]$_O$. Subjects and Objects were restricted to fantasy first names, and Verbs to present tense third person singular forms. Crucially, each sentence was paired with a contextualization paragraph, which prompted either polar question or declarative interpretation as well as narrow focus placement on either Subject, Verb, or Object, thus yielding 6 possible interpretations. The two sentences in the *Danser* set were built using similar criteria, but with no restriction on consonant voicing, with actual first names for Subjects and permitting prepositional Objects.

To exemplify, when associated with the contextualization paragraph in (1a), the sentence in (1b) was expected to be uttered as a question with narrow focus on the prepositional object:

(1) a. ‘You know your cousin Serena has relocated, probably to Lara’s, an old high-school friend of hers. You’re not sure though, so you ask your aunt:’
b. *Serena vive da Lara?*
‘Does Serena live at Lara’s?’.

Recordings were made in a sound-treated booth at Naples’ University “Federico II”, and involved 51 native speakers of the Neapolitan variety of Italian. They were all university students aged between 19 and 28 with no training in prosody, and were paid a small sum for their participation. The trials were prompted on a computer screen using Perceval (André et al. 2003), while recordings were made using an AKG MicroMic C520 head-mounted microphone connected via a Shure X2u adapter to a personal computer running Audacity (Audacity Development Team 2006). For the *Orlando* set, 30 speakers read 3 repetitions of the 6 interpretations for each of the 3 sentences, totaling 1620 items. For the *Danser* set, 21 speakers read 3 repetitions of the 6 interpretations for each of the 2 sentences, for a total of 756 utterances. A small number (ca. 3%) of the recorded utterances, containing disfluencies or prosodic breaks after the focused constituent, were excluded from the analysis.

For the evaluation of H1, we used the total duration of each individual utterance, as segmented from the parent recording session using the silence detection procedure in Praat (Boersma and Weenink 2008) followed by manual verification. Utterances were then phone segmented using Assi (Cangemi et al. 2011), a forced-alignment tool for Italian. The individual durations for about 37,000 segments were used for the evaluation of H2.

2.4 Results

Analyses were run separately on the two sets, and yielded similar results. Hence, in the following section we present results only for the *Danser* corpus.

We ran a linear mixed model predicting the dependent variable Utterance Duration by using the fixed factors Modality (question or statement), Focus (on Subject, Verb, or Object), and Sentence (two levels) and adding a random intercept for the 21 Speakers. Both the factor Modality and its interactions with the factor Focus did not reach significance ($t < 2$). A Likelihood Ratio Test compared the model with the fixed factors Focus and Sentence (and their interaction) with a model including Modality as well, and showed no significant differences ($\chi^2 = 9.9$, $df = 6$, $p = 0.13$). Hence, the hypothesis of a difference in global utterance duration across sentence modalities (H1) was rejected.

We then tested the hypothesis of localized phone duration differences (H2). We ran a linear mixed model predicting Phone Duration from three fixed factors: Focus, Sentence, and the Combination of Phone Position (from the consonant in the first syllable, C1, to the vowel in the last syllable, V8) and Modality (question or statement), adding a random intercept for the 21 Speakers. In order to verify which phone position yielded significantly different durational values across modalities, a successive difference contrast was associated to the 32 levels of the factor Combination. By pruning significant contrasts with a combined effect size of less than 10 ms or due to interactions with the Focus factor, two highly significant contrasts ($pMCMC < 0.001$) indicated that the first segment (C1) is about 12 ms longer in Statements while the last segment (V8) is about 20 ms longer in Questions. Figure 2 shows a plot of the utterance-normalized phone Durations (y-axis) for phones in each of the 16 Positions (Consonant or Vowel from 1st to 8th syllable; x-axis), pooled across Modalities (questions in dashed black line and statements in continuous gray line). The plot provides a clear visual translation of the statistical analyses: durational differences are localized at utterance boundaries, affecting the first and the last segment, and they counterbalance

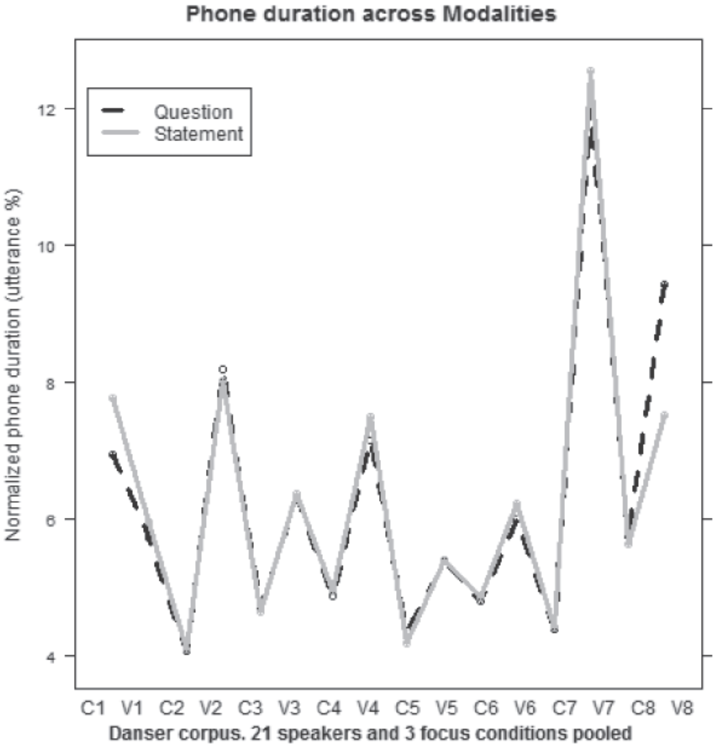

Fig. 2: Phone durations (y-axis, normalized on parent utterance duration) for the 16 phone positions (x-axis, with respect to the position of the phone within the syllable and of the syllable within the utterance) in questions and statements (dashed black and continuous gray lines, respectively) from the Danser corpus.

each other, since the first segment is longer in statements and the last is longer in questions.

2.5 Discussion

Our analysis of sentence modality contrasts in production shows that questions and statements, in addition to being intonationally specified by different tunes, are also characterized by different durational patterns. In contrast with results from the literature on different languages, our study shows that durational differences are localized at specific portions of the utterances (viz. the first and last segment) but that global utterance duration is not significantly different for questions and statements. In the following, we address the issue of whether these systematically produced durational differences are also relevant in perception, or in other words whether phonological representations for sentence modality contrasts should be enriched with respect to their temporal dimension.

3 Perception

3.1 Hypotheses

The results presented above show that Subject-Verb-Object sentences containing the same lexical material are uttered by Neapolitan Italian speakers with a different durational pattern when read with either a polar question or a declarative intonation (with narrow focus on either S, V, or O position). Specifically, questions display shorter phone durations at the beginning of the utterance, while declaratives are characterized by shorter phone durations at utterance end.

These results are compatible with the hypothesis that questions and declaratives, in addition to being differently intonationally specified (viz. by different tunes), are also phonologically contrasting along the dimension of tempo, namely through different temporal patterns. In this view, which is compatible with Lehiste's account of prosody, tempo and intonation are *orthogonal* (H1). However, if intonational contrasts are taken to be cued by all suprasegmental features, as in Ladd's account of the AM framework, differences in phone durations could also be included in intonational representations. In this view, tempo is *nested* within intonation (H2). In this case, different durational patterns could arise, for example, as a by-product of the use of different pitch accents, which would be specified by both melodic and temporal information. In this case, durational differences would be due to the phonetic implementation of intonational contrasts, and there would be no need to posit an orthogonal dimension for tempo.

Both hypotheses, however, assume that the durational differences reported in the production studies are also relevant for perception, and that they interact with F_0 movements in cueing sentence modality contrasts. That is, in case our data show that tempo needs to be included in phonological representation, we might then ask whether it should be considered as either orthogonal to (H1) or else nested within (H2) intonation. This assumption must be questioned through the evaluation of the null hypothesis stating that durational differences do not cue sentence modality contrasts (H0). Note though that our null hypothesis is challenged by the acoustic evidence discussed above, though it is consistent with both claims on the paralinguistic nature of tempo-related contrasts and, especially, with the long-term priority accorded to fundamental frequency in research on post-lexical meaning. In this case, there would be evidence for the adequacy of information reduction in abstractionist models of intonation. In very general terms:

H0: *Null hypothesis*. Durational differences do not cue sentence modality contrasts.

H1: *Orthogonality hypothesis*. Durational differences cue sentence modality contrasts and should be organized on the phonological dimension of tempo, which constitutes one of the prosodic axes, along with intonation.

H2: *Nesting hypothesis*. Durational differences cue sentence modality contrasts as part of the phonetic specification of contrasts on the phonological dimension of intonation.

An identification task with reaction times measured from stimulus offset was carried out in order to provide the relevant data (for an operationalized version of the hypotheses, see Section 3.3). If tempo can be evaluated independently from intonation (H1), we would expect identification to be affected by temporal manipulations. That is, we expect different responses to stimuli with different temporal patterns but the same intonation contour. In particular, if temporal cues are ancillary to melodic ones, we would expect the magnitude of differences in responses to temporally manipulated stimuli to be lower than that of melodically manipulated stimuli. Moreover, we could predict that the effect of temporal manipulation would increase if melodic information is made ambiguous or unavailable. If, on the other hand, phonetic temporal information is only nested within phonological intonational categories (H2), we can expect that stimuli resynthesized to have mismatching cues would still be categorized according to melodic information, but require longer processing and thus elicit longer reaction times. And if durational differences are not used in perception at all (H0), we would expect the absence of an effect of temporal manipulation on both responses and response times.

3.2 Method

Twenty-six students enrolled in non-linguistic programs at the Faculty of Arts of Naples' University "Federico II", all native speakers of the Neapolitan variety of Italian, participated in a forced-choice categorization task. Stimulus presentation and response recording were managed by Perceval (André et al. 2003); subjects were asked to use a two-button response box to code audio stimuli as either questions or statements (declaratives). Listeners were familiarized with the task and the equipment through a brief training session which included 8 non-experimental items presented via loudspeakers, then they were asked to wear headphones for the remainder of the experiment. The experimental items consisted of 18 resynthesized stimuli (see Section 3.2.1), which were created by using two utterances from the *Danser* corpus as base stimuli (Section 2.3). Given that durational differences are known to play a role in both the production and perception of focus placement in various Romance languages (see, among others, Bertinetto [1981] on Italian and Frota [2012] on Portuguese), base stimuli were the question and statement version (coded as *bQ* and *bS*, where *b* stands for "base") of the sentence (2) in its Subject-focused version alone, as read by the same female speaker.

(2) *Danilo vola da Roma*
Danilo fly-3s from Rome
'Danilo takes the Rome flight'.

The 18 experimental items were block randomized and interspersed with twice as many filler stimuli in order to avoid the use of a stereotypical intonation pattern; each block was presented 3 times to each of the 26 subjects. For each experimental trial we recorded both subjects' responses and their reaction times from stimulus offset, yielding a total of 2,808 observations.

3.2.1 Resynthesis

We used a resynthesis procedure based on work by Gubian, Cangemi, and Boves (2010, 2011) and implemented through a set of scripts in R (R Development Team 2008) and Praat (Boersma and Weenink 2008). We extracted the segmentally aligned F_0 contours of the two base stimuli (*fQ*, *fS*) and turned them into continuous functions through b-spline smoothing. That is, the F_0 curve was not discretized as is usually done in perceptual studies involving resynthesis. No top-down knowledge was fed into the resynthesis procedure, apart from anchoring the F_0 contours to the segmental boundaries. Moreover, by using continuous phonetic representations instead of a sequence of turning points, we avoided losing potentially useful melodic information. Minimalist top-down based assumptions were also made in the extraction of the durational patterns of the two base stimuli (*dQ*, *dS*), for which we stored the duration of each phone as annotated by manual segmentation.

Then we calculated an acoustically ambiguous durational pattern (*dA*), by averaging phone durations, and an acoustically ambiguous F_0 contour (*fA*), by averaging functions with respect to the segmental landmarks. Function averaging was accomplished by extracting a transform function which turned a given contour into the opposite, and by applying it with a weight of $c = 0.5$. We resynthesized each of the two base stimuli with the nine combinations between the two factors (*f* and *d*) and their three levels (*Q*, *S*, and *A*), thus obtaining 18 experimental items.

Items were coded by concatenating information about the base stimulus (*bQ* or *bS*), the F_0 contour (*fQ*, *fS*, or *fA*) and the durational pattern (*dQ*, *dS*, or *dA*). For example, the item resynthesized from a question base by keeping its original question contour but by switching to statement durational pattern was coded as *bQfQdS*. In the following, we will use *X* as an indicator of pooling: for example, *bXfQdS* indicates stimuli with question F_0 contour (*fQ*) and statement durational pattern (*dS*), resynthesized from either base (*bX*). In the graphs (and when possible in the text as well), indication of the base stimulus is dropped altogether, since no base-related effect is attested in the results (see Sections 3.4.1–2 and 3.5.2).

3.3 Predictions

If durational differences act as a cue to phonological temporal contrasts, we would expect to find significantly different responses for stimuli with different durational patterns but the same F_0 contours. That is, we expect more question responses for stimulus *fXdQ* than for stimulus *fXdS*. The bar graph in Figure 3 shows a stylization of the expected results in terms of question response percent. Items with a given F_0 contour are grouped in three triplets (*fQ*, black left; *fA*, grey central; *fS*, black right). Inside each triplet, three bars represent items with question (left, *dQ*), ambiguous (centre, *dA*), and statement (right, *dS*) durational pattern. If there is an effect of temporal manipulation, we expect question responses to decrease within triplets. Moreover, if duration acts as a secondary (compared to F_0 contour) prosodic cue to sentence modality contrasts, we would expect a stronger effect of tempo manipulation when intonation is ambiguous. That is, we expect a greater difference between *fAdQ* and *fAdS*. This is illustrated by the sharper fall in the grey triplet (*fA*), compared with the fall in the black triplets (*fQ* and *fS*); for a discussion of the quality of ambiguous intonation resynthesis in the operationalization of H1, see Section 3.5.2. These hypothesized results could be taken as support for H1.

H2, on the other hand, could be supported even in the absence of significant differences in subjects' responses, and namely by different reaction times. If phonetic durational information participates in shaping phonological intonational contrasts, listeners could still exclusively rely on melodic information for their categorization decisions, but we would expect processing to be affected by a mis-

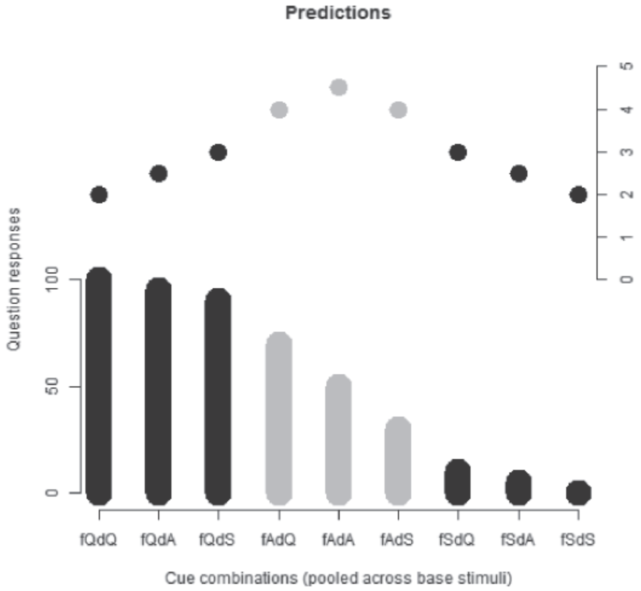

Fig. 3: Expected responses (y-axis, bar graph, Question%) and reaction times (y-axis, point graph, seconds) for the 9 resynthesis conditions (x-axis).

match between melodic and temporal information. Specifically, we would expect shorter reaction times for stimuli with congruous intonational and temporal cues (*fQdQ* and *fSdS*) compared with stimuli with incongruous information (*fQdS* and *fSdQ*). This is shown by the point graph in Figure 3, where stimuli with matching information (at both ends of the x-axis) elicit shorter reaction times than stimuli with mismatching information (close to the grey triplet). For a discussion on the use of reaction times for the operationalization of H2, see Section 3.5.3.

Absence of a significant effect of temporal manipulations on both subjects' responses and reaction times would yield instead support for H0. The two alternative hypotheses will be tested using generalized mixed logit and linear mixed models, respectively for H1 and H2.

3.4 Results

A qualitative exploration of the raw data already indicates that melodic cues alone are relevant in the perception of sentence modality contrasts (Figures 4 and 5; for the quantitative analyses, see Sections 3.4.1–2). As for identification results (Figure 4), the strong effect of melodic manipulations is attested by the drop in

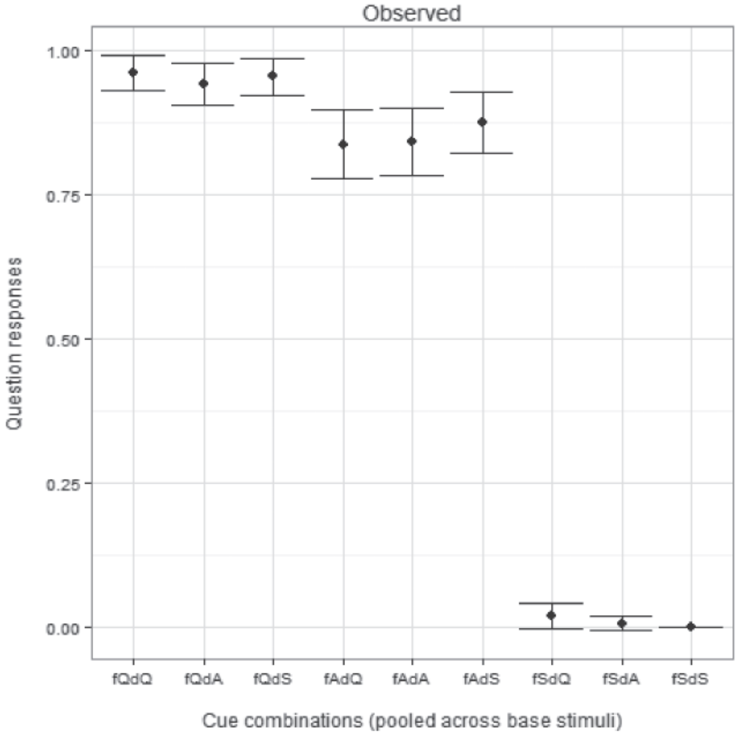

Fig. 4: Observed responses (y-axis, question%, means, and standard errors) for the 9 resynthesis conditions (x-axis).

question responses *across* triplets. Stimuli with question intonation (*fQdX*, first triplet) elicited more question responses than stimuli with ambiguous intonation (*fAdX*, second triplet), which elicited more than stimuli with statement intonation (*fSdX*, third triplet), as expected.[4] Temporal manipulations, on the other hand, did not seem to affect subjects' responses, as attested by the absence of a visible drop in question responses within triplets. For example, given stimuli with question intonation (*fQdX*, first triplet), there was no drop in response rates from stimuli with question durational pattern (*fQdQ*, first bar) to stimuli with statement durational pattern (*fQdS*, third bar).

Reaction times from stimulus offset also showed no effect of temporal manipulations (Figure 5). In particular, subjects' responses were not faster when tempo-

[4] For a discussion of the consistent bias towards question responses to stimuli with ambiguous intonation, see Section 3.5.2.

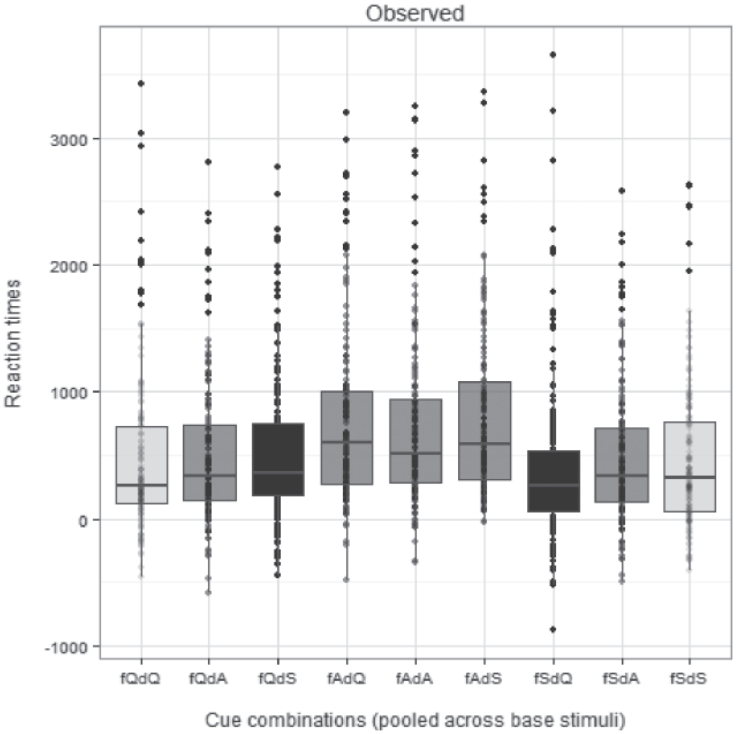

Fig. 5: Observed reaction time distributions (y-axis, ms) for the 9 resynthesis conditions (x-axis). Congruous conditions in white (fQdQ and fSdS), incongruous conditions in black (fQdS and fSdQ); for the sake of completeness, we also plot non-relevant conditions in grey.

ral and intonational cues were congruous (*fQdQ* and *fSdS*, white boxes) rather than incongruous (*fQdS* and *fSdQ*, black boxes). Only stimuli with ambiguous intonation (*fAdX*) elicited longer reaction times.

3.4.1 Orthogonality hypothesis

In order to test H1 we ran a series of generalized mixed logit models, aimed at evaluating the effect of temporal manipulations on subjects' identification responses. The most comprehensive model included the three fixed factors Intonation (three levels: question, ambiguous, and statement), Tempo (question, ambiguous, and statement), and Base Stimulus (question and statement), as well as their interactions, together with a random intercept for our 26 Subjects. We then pruned the model, excluding three-way interactions first, then two-way interactions, and ultimately non-significant factors. As a result, the comparison between the most comprehensive model and the one containing the fixed factor Intonation alone showed no significant Likelihood difference ($\chi^2 = 16.7$, $df = 15$, $p = 0.33$), thus leading to the rejection of H1.

A fortiori, the corollary of a stronger effect of temporal manipulations for stimuli with ambiguous intonation is not validated either. The corollary would have been validated by significant interactions between Intonation and Tempo. However, as can be inferred from the comparison between the model including Intonation, Tempo, Base, and their interactions with the model containing Intonation alone, no difference in Likelihood was found when comparing the model containing Intonation, Tempo, and their interaction with the model containing Intonation and Tempo but no interaction ($\chi^2 = 5.6$, $df = 4$, $p = 0$ .23).

3.4.2 Nesting hypothesis

As for H2, we ran a series of generalized linear mixed models evaluating the effect of temporal manipulations on subjects' reaction times. Prior to modeling, latencies were made positive and log-transformed. The most comprehensive model included the three fixed factors Intonation (three levels: question, ambiguous, and statement), Tempo (question, ambiguous, and statement), and Base Stimulus (question and statement), as well as their interactions, together with a random intercept for our 26 Subjects. A Likelihood Ratio Test between this model and the one without the Base factor showed no significant difference ($\chi^2 = 14.32$, $df = 9$, $p = 0.11$), so in the following we will only refer to the more economical model.

In this model, differences between response times to stimuli with congruous (*fQdQ* and *fSdS*) and incongruous (*fSdQ* and *fQdS*) sets of intonational and temporal cues are estimated by the interactions between intonation and tempo.[5] A com-

[5] Specifically, if we take latencies for *fQdQ* as a reference (*i*, intercept), we have to estimate the coefficient for temporal manipulation from Question to Statement (*tempoS*) in the case of *fQdS*, the coefficient for intonational manipulation from Question to Statement (*intonS*) in the case of *fSdQ*, and the two previously mentioned coefficients along with the interaction between temporal and durational manipulations from Questions to Statement (*tempoSintonS*) in the case of *fSdS*. Grouping the four stimuli in the two congruous and incongruous conditions, we obtain that different latencies among the two groups require a significant coefficient for *tempoSintonS*:

$$\begin{aligned}(fQdQ + fSdS) - (fSdQ + fQdS) &= (i + i + intonS + tempoS + tempoSintonS)\\ &\quad - (i + intonS + i + tempoS)\\ &= tempoSintonS\end{aligned}$$

In the model discussed in the text, this coefficient is non-significant ($t = 0.11$).

parison between the model including Intonation and Tempo as well as their interaction and the model including Intonation and Tempo with no interaction showed no significant Likelihood difference ($\chi^2 = 2.66$, $df = 4$, $p = 0.61$), thus leading to the rejection of H2.

3.5 Discussion

Our results suggest that listeners do not use durational patterns as a cue for the identification of resynthesized Neapolitan questions and statements. Specifically, listeners' responses were not affected by resynthesis of temporal patterns, and this held true also when intonational cues were made acoustically ambiguous. Our findings speak against a model of prosody in which tempo is seen as an orthogonal dimension to intonation, contra H1. Moreover, identification of sentence modality did not seem to be hindered by a mismatch between melodic and temporal cues: reaction times were in fact similar in responses to stimuli with either congruous or incongruous cues. In fact, in this case as well, listeners only seemed to rely on intonation. This finding is not consistent with the hypothesis that temporal information is part of representations for intonational categories, hence not supporting H2. Response times increased only when intonation cues were made ambiguous, which can be taken as additional evidence for the exclusive role of F_0.

In sum, systematically produced temporal detail does not seem to function as a cue to the perception of sentence modality contrasts in the Neapolitan variety of Italian. Abstraction mechanisms seem to be involved in the perception of post-lexical contrasts, in the sense that some phonetic information, even if systematically produced, does not appear to play a role in off-line identification tasks. However, to avoid overgeneralization, in the next sections we question the exact scope of our findings.

3.5.1 Design-related issues

Before discussing some epistemological issues (Section 3.6) and concluding on the possible relevance of our findings in the broader frame of research on prosodic detail (Section 4), we now turn to an examination of some issues in the experimental design which could have affected our results, and report on a small-scale spin-off experiment which addressed part of them.

3.5.2 Resynthesis

Experimental stimuli were created with an innovative resynthesis procedure, which combines modifications of F_0 contours and of temporal patterns (see Section 3.2.1). Since the procedure was elaborated for the purposes of this study, no independent testing of its performance was available. However, subjects' responses show that cross-modality manipulation was extremely successful. Identification responses were not affected by the nature of the stimulus used as base for the resynthesis: question response rates for natural questions (*bQfQdX*) and for natural statements resynthesized with question intonation (*bSfQdX*) were not significantly different, and the same holds for question to statement resynthesis.[6]

On the other hand, our resynthesis procedure could not produce a truly perceptually ambiguous F_0 contour between questions and statements. Stimuli intended to be intonationally ambiguous (*fAtX*) were identified as questions well above chance level (see Figure 4). This finding is not surprising. As noted above (see Section 3.2.1), ambiguous contours were obtained by setting *c*, the weight of the transform function, to 0.5. In other words, ambiguous contours were calculated as the mean of the two time-warped F_0 contours, thus qualifying as merely *acoustically* ambiguous contours. These stimuli would also be *perceptually* ambiguous only if the perceptual space between questions and statements were linear. However, the non-linear warping of perceptual space has been long acknowledged for contrasts at the segmental level, and this makes the assumption of a linear perceptual space for utterance-wide intonational contrasts even more untenable (Gubian, Cangemi, and Boves 2010). By allowing a fine-grained control of separate acoustic dimensions, our resynthesis procedure is indeed especially suited for the exploration of the perceptual space of intonational contrasts. This would ultimately provide the necessary insights for the creation of truly perceptually ambiguous stimuli. However, in the absence of such preparatory work, for the purposes of our current study, we consciously restrained to the simplistic choice of using acoustically ambiguous stimuli.

We operationalized the hypothesis of an orthogonal processing of tempo and intonation (H1) by predicting an effect of temporal manipulation within triplets,

[6] In the generalized mixed logit model predicting subjects' responses from Intonation and Base as fixed factors, together with their interactions and adding a random intercept for Subjects, no base-related coefficient reaches the significance level. Excluding the *Base:Intonation* interaction for stimuli with ambiguous intonation (which are not relevant for the present discussion on cross-modality resynthesis, and in any case non-significant as well: $\beta = 0.98$, $z = 1.70$, $p = 0.09$), we find non-significant *Base* ($\beta = -0.45$, $z = -0.93$, $p = 0.35$) and *Base:Intonation* interaction ($\beta = 0.45$, $z = 0.35$, $p = 0.72$).

and by extracting the corollary prediction of stronger effects in the ambiguous intonation triplet (see Section 3.3). Our results show no within-triplet effect, thus leading to the rejection of H1, independent of the quality of ambiguous stimuli resynthesis. However, H1 could also have been operationalized in a more restrictive way, namely by predicting an effect of temporal manipulation only when intonational cues are truly perceptually ambiguous. The preparatory work on the warping of perceptual space for utterance-wide intonational contrasts required for the evaluation of this stricter hypothesis is ongoing.

3.5.3 Reaction times

As we said above (see Section 3.3), the evaluation of H2 needs particular caution. We suggested that the relevance of temporal information in the phonological representations of intonational contrasts (nesting hypothesis) could have been indicated by different reaction times between stimuli with congruous (*fQdQ* and *fSdS*) and incongruous (*fSdQ* and *fSdQ*) information on the melodic and temporal levels. However, the quantitative evaluation of significant differences in reaction times is affected by two kinds of issues.

On a lower level, a first problem is represented by the fact that total stimulus duration was not fixed. We showed at Section 2.4 that global utterance duration in Neapolitan Italian is not significantly different across questions and statements. However, this does not mean that differences in the duration of individual items are not attested at all. Specifically, the durations of the utterances used as bases for the resynthesis procedure were 1.2 s for the statement version (*fXdS*) and 1.3 s for the question (*fXdQ*). Since resynthesis of ambiguous durational patterns was based on pattern means (see Section 3.2.1), duration of temporally ambiguous stimuli was 1.25 s (*fXdA*). Given that Perceval calculates reaction times with reference to stimulus onset, stimulus duration was subtracted from latencies in order to obtain results which can be interpreted as reaction times from the end of the stimulus.

Hence, given the nature of our stimuli, one might wonder whether reaction times can be really considered as a reliable measure for the validation of H2. Response latencies from stimulus end would be an indicator of ease of processing only if listeners delayed the evaluation of (mis)match between melodic and temporal information until the end of stimulus. This, however, is only a simplifying assumption, made in the absence of relevant knowledge on the integration of suprasegmental cues in perception of post-lexical meaning. Indeed, there is reason to believe that, since temporal and melodic cues unfold in time, their integration could be best captured by on-line tasks or by the monitoring of multiple references for reaction times. Specifically, as for intonation, the very idea of analyzing phonetic data (F_0 contours) as a succession of phonological events (pitch accents and edge tones) entails the existence of multiple points in time where bundles of perceptually relevant material are made available. Future studies on the interplay between temporal and melodic cues should definitely take into account the possibility of evaluating cue integration as time unfolds. On-line tasks might prove especially useful in this sense, as shown by recent work on the use of semantic scaling in the perception of focus placement in European Portuguese (Frota 2012) and on indirect identification for sentence modality contrast in Northern Standard German (Petrone and Niebuhr in press).[7]

As for the more restricted purposes of our study, given that intonational (and possible temporal) cues unfold in time, the use of response latencies relative to utterance end is a factor which could have limited the conclusiveness of H2 validation. If differences in reaction times to congruous and incongruous stimuli had been found to be statistically significant, it would have been difficult to evaluate their meaningfulness. However, since our data reveal no significant differences, this point does not need further elaboration.

3.5.4 Spin-off

As we said above, if we assume that tempo is a secondary cue to sentence modality, a corollary of the orthogonality hypothesis (H1) is that differences in durational patterns are best perceived when intonational cues are not available. That is, we expect a greater difference between responses to stimuli with question and statement durational patterns for items with ambiguous F_0 contours (*fA*) than for items with either question or statement F_0 contour (*fQ* or *fS*). This motivates the steeper fall in the hypothesized rate of question responses for the grey triplet in Figure 3, compared to the black triplets. That is, maximal difference in subjects' responses is expected between the pair *fAdQ* and *fAdS*.

Our results did not support the orthogonality hypothesis (see Section 3.4.1). However, given the set of stimuli used in the task, we cannot rule out the possibility of a ceiling effect in subjects' responses due to the availability of intonationally unambiguous stimuli. That is, listeners' attention might have been diverted from subtle temporal cues because of the presence of striking intonational differ-

[7] Sentence modality contrasts are not especially well suited for eye-tracking studies, but if listeners were asked to provide graded acceptability judgments rather than forced-choice responses, eye movements along the scale during stimulus presentation could perhaps prove useful.

ences; as Hawkins (2011) puts it, "listeners seem to learn about new phonetic detail when it does not contradict other important cues to communicating meaning." For this reason, we devised a short spin-off experiment to be run after the main test. Subjects were asked to identify as question or statements stimuli in the *fAdQ* and *fAdS* conditions alone; no fillers or intonationally clear stimuli were presented. We hypothesized that, if durational differences are perceptible and used in sentence modality categorization, presenting intonationally ambiguous stimuli alone would maximize the visibility of the effect of tempo on perception. We gathered responses from the same 26 subjects for the 2 conditions resynthesized from 2 bases, presented 4 times in each of 2 independently randomized blocks, for a total of 832 items. We predicted subjects' Responses using a generalized linear mixed model with Tempo as a fixed factor and a random intercept for Subjects. The coefficient for Tempo, however, did not prove significant ($\beta = -0.03$, $z = -0.199$, $p = 0.843$). This indicates that, even when stimuli are presented so as to maximize listeners' attention to temporal manipulations, durational differences are not a cue to sentence modality contrasts. However, as in the evaluation of H1, given that resynthesized intonation was only acoustically (rather than perceptually) ambiguous, the evidence gathered cannot be considered as truly conclusive.

3.6 Epistemological issues

The main limitation to the use of our results in drawing clear-cut conclusions on the role of tempo as a prosodic detail, however, comes from a different source. Our experiment aimed at evaluating whether the acoustic differences in durational patterns we documented in Section 2 above are used as a cue in the perception of sentence modality contrasts. A positive answer to this research question would have implied that tempo has to be somehow included in phonological representations of intonation or prosody. The negative evidence (H0) we gathered through our experiment, on the other hand, does not allow us to draw the opposite conclusion, namely that tempo should not be included in phonological representations. The scope of generalization for negative findings must be accurately determined. We can only state that off-line judgments on the perception of sentence modality contrasts in clean read speech in Neapolitan Italian are not affected by durational differences, but we cannot rule out that durational differences play a role in the perception of other linguistically structured contrasts, in other communicative contexts, or in other languages.

To be more specific, the scope of our results can be further narrowed down to the conclusion that perception of sentence modality contrasts is not affected by the durational differences contained in our two base stimuli. In a radical perspective, we cannot exclude that choosing a different pair of base stimuli could have affected our results. The two base stimuli were chosen on the basis of results from the production study (see Section 2.4): among sentences with the same lexical material and read by the same speaker, we chose the two utterances with best fit to the durational patterns which characterized statements and questions in the two corpora used in the production studies cited above. However, this does not mean that the perceptual evaluation in this study only relates to the modeling we previously provided for production data. As we said above (see Section 3.2.1), we used a data-driven resynthesis procedure, in which top-down assumptions were as minimalist as possible. As a result, the downfalls of an incorrect modeling of production data are strongly limited. For example, our intonational resynthesis gave excellent results (see Section 3.5.2) by simply warping time-aligned F_0 contours. In other words, all previous knowledge from the literature on the phonetics and phonology of intonation was condensed and limited to 'alignment of F_0 contours with segmental boundaries is relevant.' For cross-modality resynthesis of utterances with the same segmental content, this top-down information alone allowed for a 94% shift[8] in subjects' identification responses. For temporal resynthesis, we limited our assumptions to 'variations in segmental durations are relevant,' as the most general formulation of the findings from our production studies. The segmental level was preferred over smaller (i.e., subsegmental phases) or larger (e.g., syllables) domains, as a reasonable compromise between a fine-grained temporal analysis and the degree of precision allowed by Assi, the forced alignment tool used in the production studies (see Section 2.3). With such minimalist assumptions on both dimensions and with such clearly opposite effects on subjects' responses, it is reasonable to infer that F_0 contours and durational patterns play very different roles in sentence modality perception.

In sum, as in the case of every study in which the alternative hypotheses are not supported, no general statement can be inferred from our results. We have to limit the scope of our conclusions according to the features of our study, and restrain from claiming that temporal detail is irrelevant in cueing post-lexical meaning. Moreover, as discussed above, we cannot rule out the possibility that on-line tasks featuring truly perceptually ambiguous stimuli are necessary if we want to meet the standard of "good effort" (in the terms of Frick 1995) required to

[8] This is the absolute value of the estimated coefficient of question to statement manipulation in a linear mixed model predicting subjects' Response (coded as a continuous variable) from Intonation as a fixed factor with a random intercept for Subject. The coefficient is highly significant ($t = 62.53$).

accept the null hypothesis, according to which phonological representations of intonation do not need to be enriched with temporal detail.

In our opinion, we can nonetheless conclude that its effects are hard to track in sentence modality contrasts in Neapolitan Italian read speech, that its exploration is unlikely to reveal the need for an enrichment of the phonological structure of prosody, and that its perceptual evaluation is ultimately a sorely unrewarding enterprise.

4 Conclusion

In the first part of this study, we showed that the production of questions and statements in Neapolitan Italian is characterized by subtle but consistent differences in segmental durations. However, these acoustic differences do not seem to be used as perceptual cues: listeners' responses in a forced choice identification task were not affected by the manipulation of durational patterns. Moreover, no difference was found in response times to stimuli with congruous and incongruous information on the temporal and on the melodic levels. These findings are not incompatible with an abstractionist view of perception of post-lexical contrasts, in which not all of the systematically produced phonetic information appears to play a clear role in perception. However, both methodological and epistemological issues prevent us from considering the evidence gathered in this study as truly conclusive. On the one side, the multiparametric resynthesis procedure used in the creation of the experimental stimuli should be refined, especially as far as the creation of intonationally ambiguous stimuli is concerned. Our procedure performs nonetheless very well in cross-modal resynthesis of intonation, and could represent a useful tool in the exploration of perceptual space at the utterance level. On the epistemological side, since our experiment was designed to test the perceptual importance of temporal detail, we only found evidence supporting the null hypothesis, a fact which limits the generalizability of our findings.

Acknowledgments

We would like to thank the editors and two anonymous reviewers for constructive criticism, Robert Espesser for support in the statistical analyses, the Dipartimento di Filologia Moderna of Naples' University "Federico II" for providing the infrastructure (sound-treated booth and silent rooms) for conducting the experiments, and Alan Yu and the audience of LabPhon 13 for useful suggestions.

References

André, Carine, Alain Ghio, Christian Cavé, & Bernard Teston. 2003. Perceval: a computer-driven system for experimentation on auditory and visual perception. In Maria-Josep Solé, Daniel Recasens, & Joaquín Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 1421–1424. Barcelona: Futurgraphic.

Audacity Development Team. 2006. Audacity [computer program]. http://www.sourceforge.net

Bertinetto, Pier Marco. 1981. *Strutture prosodiche dell'Italiano: accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Academia della Crusca

Boersma, Paul, & David Weenink. 2008. Praat: doing phonetics by computer [computer program]. http://praat.org/.

Cangemi, Francesco, Francesco Cutugno, Bogdan Ludusan, Dino Seppi, & Dirk Van Compernolle. 2011. Automatic speech segmentation for Italian (ASSI): tools, models, evaluation and application. In Barbara Gili Fivela, Antonio Stella, Luigia Garrapa, & Mirko Grimaldi (eds.), *Contesto comunicativo e variabilità nella produzione e percezione della lingua. Proceedings of the 7th Associazione Italiana di Scienze della Voce Conference*. Roma: Bulzoni.

Cangemi, Francesco, & Mariapaola D'Imperio. 2011a. Local speech rate differences between questions and statements in Italian. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 392–395. Hong Kong: City University of Hong Kong.

Cangemi, Francesco, & Mariapaola D'Imperio. 2011b. Prosodia oltre la f0: Tempo e modalità. In Barbara Gili Fivela, Antonio Stella, Luigia Garrapa, & Mirko Grimaldi (eds.), *Contesto comunicativo e variabilità nella produzione e percezione della lingua. Proceedings of the 7th Associazione Italiana di Scienze della Voce Conference*. Roma: Bulzoni.

Cangemi, Francesco, & Mariapaola D'Imperio. To appear. Sentence modality and tempo in Neapolitan Italian. In Joaquín Romero & María Riera (eds.), *Selected Papers from the 5th Phonetics and Phonology in Iberia Conference*. Amsterdam: John Benjamins.

De Dominicis, Amedeo. 2010. Interrogative e assertive in un corpus dialettale recuperato (Bomarzo). In Franco Cutugno, Pietro Maturi, Renata Savy, Giovanni Abete, & Iolanda Alfano (eds.), *Parlare con le persone, parlare alle macchine: La dimensione interazionale della comunicazione verbale. Proceedings of the 6th Associazione Italiana di Scienze della Voce Conference*. Torriana: EDK.

D'Imperio, Mariapaola. 2001. Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33(4). 339–356.

D'Imperio, Mariapaola, & Francesco Cangemi. 2011. Phrasing, register level downstep and partial topic constructions in Neapolitan Italian. In Christoph Gabriel & Conxita Lleó (eds.), *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies*, 75–94. Amsterdam: John Benjamins.

D'Imperio, Mariapaola, & David House. 1997. Perception of questions and statements in Neapolitan Italian. In Georgios Kokkinakis, Nikos Fakotakis, & Evaggelos Dermatas (eds.), *Proceedings of the 5th European Conference on Speech Communication and Technology*, 251–254. Rhodes.

Dryer, Matthew S. 2011. Polar questions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/chapter/116

Fagyal, Zsuzsanna. 1997. Chanting intonation in French. *University of Pennsylvania Working Papers in Linguistics* 4(2). 77–90.

Frick, Robert. 1995. Accepting the null hypothesis. *Memory & Cognition* 23. 132–138.

Frota, Sónia. 2012. A focus intonational morpheme in European Portuguese: Production and perception. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and Meaning*, 163–196. Berlin/New York: Mouton de Gruyter.

Frota, Sónia, Marisa Cruz, Flaviane Svartman, Marina Vigário, Gisela Collischonn, Aline Fonseca, & Carolina Serra. To appear. Intonational variation in Portuguese: European and Brazilian varieties. In Sónia Frota & Pilar Prieto (eds.), *Intonational variation in Romance*. Oxford: Oxford University Press.

Goldinger, Stephen. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5). 1166–1183.

Grice, Martine, Mariapaola D'Imperio, Michela Savino, & Cinzia Avesani. 2005. Towards a strategy for labeling varieties of Italian. In Sun-Ah Jun (ed.), *Prosodic Typology and Transcription: A Unified Approach*, 55–83. Oxford: Oxford University Press.

Gubian, Michele, Francesco Cangemi, & Lou Boves. 2010. Automatic and data driven pitch contour manipulation with functional data analysis. In Mark Hasegawa-Johnson, Ann Bradlow, Jennifer Cole, Karen Livescu, Janet Pierrehumbert, & Chilin Shih (eds.), *Proceedings of 5th International Conference on Speech Prosody*. Chicago.

Gubian, Michele, Francesco Cangemi, & Lou Boves. 2011. Joint analysis of f0 and speech rate with functional data analysis. In *Proceedings of 36th International Conference of Acoustics, Speech and Signal Processing*, 4972–4975. Prague.

Hawkins, Sarah. 2011. Does phonetic detail guide situation-specific speech recognition? In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 9–18. Hong Kong: City University of Hong Kong.

Hayes, Bruce, & Aditi Lahiri. 1991. Durationally specified intonation in English and Bengali. In Rolf Carlson, Lennart Nord, & Johan Sundberg (eds.), *Proceedings of the 1990 Wenner-Gren Center Conference on Music, Language, Speech and the Brain*, 78–91. Houndmills/London: Macmillan.

Henriksen, Nicholas. 2012. The intonation and signaling of declarative questions in Manchego Peninsular Spanish. *Language and Speech* 55(4). 543–576.

Hirschberg, Julia, & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20. 241–251.

Johnson, Keith. 1997. Speech perception without speaker normalization. In Keith Johnson & John Mullennix (eds.), *Talker Variability in Speech Processing*, 145–165. San Diego: Academic Press.

Kohler, Klaus. 1991. A model of German intonation. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)* 25. 295–360.

Ladd, D. Robert. 2008. *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press. Original edition, Cambridge: Cambridge University Press, 1996.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge: Cambridge University Press.

Levelt, Willem. 1989. *Speaking. From Intention to Articulation*. Cambridge: Massachussets Institute of Technology Press.

Liberman, Mark. 1975. The intonational system of English. Ph.D. dissertation, Massachussets Institute of Technology.

Maturi, Pietro. 1988. L'intonazione delle frasi dichiarative ed interrogative nella varietà napoletana dell'Italiano. *Rivista Italiana di Acustica* 12. 13–30.

Muñiz Cachón, Carmen, Ruth González Rodríguez, Liliana Díaz Gómez, Mercedes Alvarellos Pedrero. 2012. Prosodia gallego-asturiana en enunciaos SVO. *Revista de Filoloxía asturiana* 6. 335–349.

Norris, Dennis. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52. 189–234.

Petrone, Caterina. 2008. Le rôle de la variabilité phonétique dans la représentation des contours intonatifs et de leur sens. Ph.D. dissertation, Université de Provence.

Petrone, Caterina, & Oliver Niebuhr. In press. On the intonation in German intonation questions: The role of the prenuclear region. *Language and Speech*.

Pierrehumbert, Janet. 1980. The phonology and phonetics of English intonation. Ph.D. dissertation, Massachussets Institut of Technology.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 137–157. Amsterdam: Benjamins.

Pierrehumbert, Janet, & Mary Beckman. 1988. *Japanese Tone Structure*. Cambridge: Massachusetts Institute of Technology Press.

R Development Team. 2008. R: A language and environment for statistical computing [Computer software manual]. Vienna: R Foundation for Statistical Computing.

Rialland, Annie. 2007. Question prosody: an African perspective. In Tomas Riad & Carlos Gussenoven (eds.), *Tones and Tunes: Studies in Word and Sentence Prosody*, 35–62. Berlin: de Gruyter.

Ryalls, John, Guylaine Le Dorze, Nathalie Lever, Lisa Ouellet, & Céline Larfeuil. 1994. The effects of age and sex on speech intonation and duration for matched statements and questions in French. *The Journal of the Acoustical Society of America* 95(4). 2274–2276.

Schweitzer, Katrin, Michael Walsh, Sasha Calhoun, & Hinrich Schütze. 2011. Prosodic variability in lexical sequences: Intonation entrenches too. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 1778–1781. Hong Kong: City University of Hong Kong.

Smith, Caroline L. 2002. Prosodic finality and sentence type in French. *Language and Speech* 45(2). 141–178.

Stevens, Kenneth N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America* 111(4). 1872–1891.

van Heuven, Vincent J., & Ellen van Zanten. 2005. Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication* 47. 87–99.

Vanrell, Maria del Mar. 2011. The phonological relevance of tonal scaling in the intonational grammar of Catalan. Ph.D. dissertation, Universitat Autònoma de Barcelona.

Ward, Gregory, & Julia Hirschberg. 1988. Intonation and propositional attitude: the pragmatics of L*+H L H%. In Joyce Powers & Kenneth de Jong (eds.), *Proceedings of the 5th Eastern States Conference on Linguistics*, 512–522. Columbus: Ohio State University Press.

Zadeh, Vaideh Abolhasani, Carlos Gussenhoven, & Mahmood Bijankhan. 2011. A pitch accent position contrast in Persian. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 188–191. Hong Kong: City University of Hong Kong.