LOCAL SPEECH RATE DIFFERENCES BETWEEN QUESTIONS AND STATEMENTS IN ITALIAN

Francesco Cangemi & Mariapaola D'Imperio

Laboratoire Parole et Langage, Université de Provence, France
Francesco.cangemi@lpl-aix.fr; mariapaola.dimperio@lpl-aix.fr

ABSTRACT

In this paper we address the issue of whether modality can be coded by cues other than pitch accent category in Neapolitan Italian. Specifically, our findings show that segmentally identical sentences uttered as either a yes/no question or a statement show different patterns of local speech rate. Specifically, while global utterance duration is the same in the two modalities, differences were found for individual phone duration, and mainly at utterance edges. These results are not compatible with a universal view of global rate differences between questions and statements and call for a more complex model of the interaction between segmental and suprasegmental contrasts.

Keywords: prosody, modality, information structure, tempo, Neapolitan Italian

1. INTRODUCTION

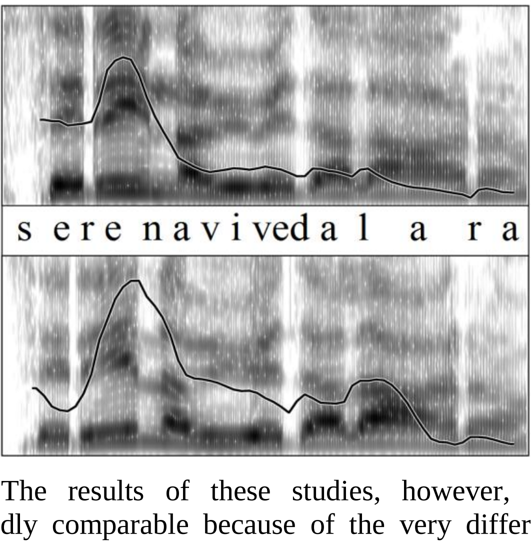
Among the prosodic cues to intonation meaning (f_0, duration, intensity, voice quality), experimental research on prosody has traditionally focused on f_0. Studies on the role of temporal structure in the production and perception of various languages rarely tend to put tempo in direct relationship with the core modules of the form-function duality in language. For example, various studies have underlined the importance of speech rate as a segmental phonetic factor (in the study of phone durations, [16]), as an idiosyncratic feature (potentially useful for speaker verification applications, [9]), as a cue to emotional speech ([17]) or as a resource for turn management ([5]).

Yet, recent studies [4] investigated the identification of intonation meaning by using stimuli with resynthesized f_0 contours superimposed on the original temporal pattern. They show that the (unmodified) temporal pattern of the base stimulus may interfere with the (modified) f_0 contour. These facts point to the necessity of a deeper understanding of the relationships between tempo and cues to intonational contrast.

1.1. Tempo and intonational meaning

A number of production studies have addressed more or less directly the issue of the relationship between tempo and intonational meaning. Speech rate patterns have been examined in direct relation to the informational value (given vs. new) and prominence pattern (accented vs. unaccented) in Dutch [7], or to sentence modality (question vs. statement) in Dutch, Manado Malay and Orkney English [10]. While mainly focusing on other prosodic cues, studies on French [15] and Neapolitan Italian have also provided evidence for different tempo patterns across sentence modality [12] and focus pattern [8]. Neapolitan Italian is especially well-suited for this kind of investigation, because sentence modality is exclusively conveyed through intonation contrast, while not employing morpho-syntactic differences (see Fig. 1).

Figure 1: Sentence '[Serena]$_F$ lives at Lara's' uttered as a Statement (top) and as a Question (bottom). Ranges: time: 0-1s, f_0: 170-320Hz.



The results of these studies, however, are hardly comparable because of the very different methodologies employed to test different claims on different languages. At the utterance level, questions appear to be globally shorter than statements in Dutch, Orkney English and Manado Malay, but the opposite appears to be true for

Neapolitan Italian. [10] also showed that tempo differences can be localized at specific portions of the utterance. As for Manado Malay, the strongest differences in duration can be found in the last foot of the utterance, while in Dutch the durational differences seem to be strongest in the mid portion of the utterance (i.e., in the stretch between the stressed syllables of Subject and Object within SOV sentences).

The results for Neapolitan Italian, although not immediately comparable to the ones cited above, also support the view that sentences uttered with different modalities not only have a different global duration, but that these differences are due to local tempo differences [12] p.83-6,162-3, which appear to be localized within the stressed vowels of the first and the last prosodic word. These results suggest that, across sentence modalities, the distribution of speech rate may not be uniform within the utterance. Our study aims at providing a detailed account of how sentence modality affects tempo in Neapolitan Italian. Since we also know that focus affects segmental duration in a variety of languages [6], the interaction between focus structure and modality was investigated as well.

1.2. Hypotheses

Note that the studies cited above found opposite effects of modality on tempo at the utterance level: questions are shorter in Manado Malay, Orkney English and Dutch [10], but longer in Neapolitan Italian [12]. Thus, our Hypothesis 1 was that *sentences have a different duration when uttered as questions or statements*, independent of the direction of the effect. H1: $U_Q \neq U_S$.

Moreover, [10, 12] found that tempo variations can be localized at specific portions of the utterance. This means that, should H1 be confirmed, it is possible that durational differences at the utterance level would not be ascribed to a uniform stretching (or compression) of individual phone durations. Specifically, we hypothesize that questions and statements might be characterized by different speech rate patterns across the utterance, rather than by a mere difference in global speech rate. Note that, in the studies cited above, differences in tempo patterns appear to be localized on units which are phonologically very different in nature: prosodic words [12], stressed syllables [12, 15], feet [10] (Malay), unstressed syllables [10] (Dutch). Since the individuation of the exact phonological domain of tempo variation goes clearly beyond the scope of this paper, we decided to select the segment as the relevant unit for analysis.

In other words, Hypothesis 2 was that *temporal differences between sentences uttered as questions or statements are not due to linear transforms of phone durations*. H2: $P\{1,n\}_Q \neq aP\{1,n\}_S$.

Table 1: Summary of Hypotheses.

Hypothesis		H2	
Confirmed		No	Yes
H1	No	NO effect	LOCAL Effect
	Yes	GLOBAL effect	GLOBAL + LOCAL eff.

Note that questions and statements could show a different speech rate pattern across the utterance, even if H1 is disconfirmed. This is because we cannot exclude *a priori* the possibility of generating the same utterance duration from two different (but counter-balanced) speech rate patterns. In this case, the verification of Hypothesis 2 would entail the existence of local tempo variations across modalities. Thus, H1 and H2 can be combined to test whether sentence modality affects tempo patterns at a global (utterance) or at a local (here, segment) level (see Table 1). Should H1 and H2 be disconfirmed, we would have evidence of a total absence of sentence modality effect on tempo patterns (H0). If H1 is confirmed and H2 is disconfirmed, we could conclude that modality affects utterance duration as a whole, Should H1 be disconfirmed and H2 confirmed, we would have evidence of local tempo variations which do not affect total utterance duration (i.e. they would be counterbalanced). In case H1 and H2 are confirmed, we could conclude that modality affects in the first place some specific portions of the utterance, and that this effect is visible in terms of total utterance duration as well.

2. MATERIAL AND METHOD

21 native speakers of Neapolitan Italian performed a reading task in a sound-treated booth. For each trial, they were asked to silently read a contextualization paragraph, then to read aloud a target sentence (in boldface). Target sentences had similar syntactic structure (NP VP PP) and number of words; words at the same syntactic position had the same number of syllables, syllabic structure (all CV, and thus the same number of phonemes as well), metrical structure (penultimate stress) and

lexical frequency. The contextualization paragraph was meant to induce one of the six possible combinations between the three-level *Focus* factor (Subject, Verb or indirect Object) and the two-level *Modality* factor (Question or Statement). We recorded a total of 2 sentences * 3 repetitions * 6 contexts * 21 speakers = 756 utterances. 45 items did not undergo analysis because they contained pauses or disfluencies.

The utterances were designed to be phone-segmented using a tool for forced alignment of Italian speech [3]. In order to achieve reliable alignment results, in the construction of the target sentences we avoided consonant clusters, glides and phones which rarely (n<6000) occurred in the dataset used to train the alignment tool.

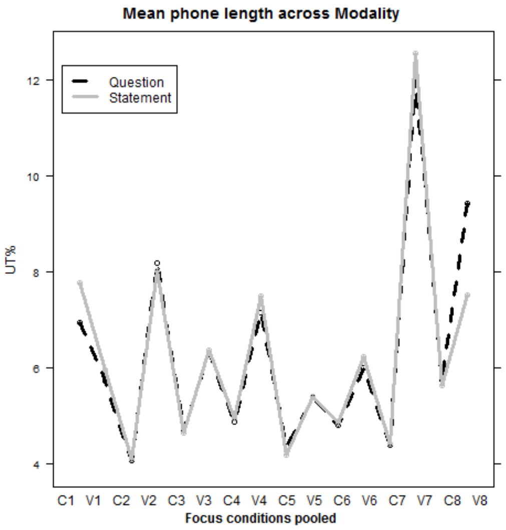
While the verification of Hypothesis 1 was based on the duration of the entire utterances, in order to test Hypothesis 2 we extracted the durations of each of the 16 phones composing a given target utterance and normalized them on utterance duration.

3. RESULTS

In order to test Hypothesis 1 we ran a linear mixed model which predicted the dependent variable *Utterance Duration* by using the fixed factors *Modality* (question or statement), *Focus* (on NP, VP or PP) and *Sentence* (two levels), adding a random intercept for the 21 *Speaker*s. Both the factor *Modality* and its interactions with the factor *Focus* did not reach significance (t<2), leading to the rejection of Hypothesis 1. A Likelihood Ratio Test comparing the model with the fixed factors *Focus* and *Sentence* (and their interaction) with a model including *Modality* as well showed no significant differences (χ^2=9.9, df=6, p=0.13). For the sake of completeness, we report that the factor *Focus* on its own did reach significance (|t|>3), indicating that, compared to NP-focused utterances, VP-focused utterances are longer while PP-focused utterances are shorter (mean difference of about ±30ms on a mean duration of 1.15s).

We then tested Hypothesis 2 by running a linear mixed model predicting phone duration from three fixed factors: *Focus* (three levels: NP, VP and PP), *Sentence* (two levels) and the *Combination* of Phone position (from C1 to V8) and Modality (Question or Statement). A successive difference contrast was associated to the 32 levels of the factor *Combination* in order to verify which phone position yielded significantly different durational values across modality. 11376 phone durations were analyzed, and a random intercept was added to account for variability across the 21 *Speakers*. A Likelihood Ratio Test showed that, compared with the model including three-way interactions, a two-way interactions model had a slightly (and significantly) smaller Likelihood, but better AIC and BIC. Consequently, in what follows we will only refer to the more economical model. Our model showed a number of significant contrasts, but since their combined size effect was less than 10ms, they will not be further commented here. Apart from that, three significant interaction coefficients between *Combinations* and *Focus* were found, indicating (together with the non-interacting contrasts) that the stressed vowel of a focused phrase is significantly longer (~10ms) in Statements. Most importantly, the two highly significant contrasts (pMCMC<0.001) indicated that the first segment (C1) is longer (~12ms) in Statements and the last segment (V8) is longer (~20ms) in Questions.

Figure 2: Phone position against duration in the two modality conditions.



A more readable account of these results can be provided by plotting, for every phone, its *Position* in the utterance (x-axis) against its *Duration* (normalized on utterance length, y-axis) for the two levels of the factor *Modality* (see Figure 2). If Modality is not significant, we expect two exactly overlapping lines; in the case of an utterance-level effect alone, on the other hand, we expect an offset between the two lines. Our results only show localized differences (mainly on C1 and V8), so we can conclude that Hypothesis 2 is confirmed.

4. DISCUSSION

The results presented in the preceding section do not support Hypothesis 1 while supporting Hypothesis 2, thus pointing to the existence of local and not global rate effects between questions and statements (see Table 1). This means that sentence modality does not affect utterance duration as a whole, but rather some specific portions (here examined at the phone level), and moreover in a way that utterance-level durational differences are neutralized. Our results are not in line with the picture emerging from previous studies on this topic, which found that questions and statements are characterized by different utterance duration.

Speculating on their results, [10] propose that the higher speech rate found in questions could be interpreted as a prosodic universal. From an ethological perspective, this effect could be taken as the temporal counterpart of high pitch values in signaling submissiveness [11], since “small (harmless) creatures have higher pitches, and make faster movements, than large (dangerous) creatures” [10] p.97. From a different perspective, building on the dichotomy between statement/relaxation (low, falling pitch) and question/tension (high, rising pitch) [2], faster speech rate in questions could also have been motivated by the fact that “high rate and acceleration go together with tension” [10] p.97. Our results do not seem to support this view, in that questions and statements were not found to have a different global speech rate, but rather a different distribution of speech rate values within the utterance.

5. CONCLUSIONS

Our study suggests that sentence modality contrasts (i.e. question vs. statement) are implemented with different tempo patterns in Neapolitan Italian. Specifically, durational differences appear to be localized at the utterance edges. These results are not compatible with universal theories linking high or rising pitch and fast speaking rate with submissive attitude and question modality.

6. ACKNOWLEDGMENTS

This work was supported by a Marie Curie grant (RTN Sound to Sense). We would like to thank the staff at CIRASS (University “Federico II”, Naples) for providing recording facilities and Robert Espesser for support in data analysis.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2011. Praat: Doing phonetics by computer. *http://www.praat.org*

[2] Bolinger, D.W. 1964. Intonation as a universal. *Proc. 9th International Congress of Linguists* Cambridge, Mass., 833-848.

[3] Cangemi, F., Cutugno, F., Ludusan, B., Seppi, D., van Compernolle, D. Forthcoming. ASSI: Automatic Speech Segmentation for Italian. *Proc. 7th Conference of Associazione Italiana di Scienze della Voce* Lecce.

[4] D'Imperio, M., Cangemi, F. 2009. The interplay between tonal alignment and rise shape in the perception of two Neapolitan rising accents. *Talk at 4th Conference on Phonetics and Phonology in Iberia* Las Palmas de Gran Canaria.

[5] Duncan, S. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology* 23(2), 283-292.

[6] Eady, S., Cooper, W., Kloouda, G., Mueller, P., Lotts, D. 1986. Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29, 233-251.

[7] Eefting, W. 1991. The effect of information value and accentuation on the duration of Dutch words, syllables and segments. *Journal of the Acoustical Society of America* 89(1), 412-424.

[8] Gubian, M., Cangemi, F., Boves, L. Forthcoming. Joint analysis of F0 and speech rate with FDA. *Proc. 36th I^{nt}ernational Conference on Acoustics, Speech and Signal Processing* Praha

[9] van Heerden, C.J., Barnard, E. 2007. Speech rate normalization used to improve speaker verification. *SAIEE Africa Research Journal* 98(4), 129-135.

[10] van Heuven, V., van Zanten, E. 2005. Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication* 47, 87-99.

[11] Ohala, J. 1984. An ethological perspective on common cross language utilization of F0 of voice. *Phonetica* 41, 1-16.

[12] Petrone, C. 2008. *Le Rôle de la Variabilité Phonétique Dans la Représentation des Contours Intonatifs et de Leur Sens.* Ph.D. Thesis, Université Aix-Marseille I.

[13] Pfitzinger, H. 2001. Phonetische analyse der sprechgeschwindigkeit. *Forschungsberichte des Instituts für Phonetik.* und Sprachliche Kommunikation der Universität München, 117-264.

[14] R Development Core Team 2010. R: A Language and Environment for Statistical Computing. *http://www.R-project.org*

[15] Smith, C.L. 2002. Prosodic finality and sentence type in French. *Language and Speech* 45(2), 141-178.

[16] Turk, A., Nakai, S., Sugahara, M. 2006. Acoustic segment durations in prosodic research: A practical guide. In Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., Schließer, J. (eds.), *Methods in Empirical Prosody Research*. Berlin, New York: de Gruyter, 1-28.

[17] Williams, C., Stevens K. 1972. Emotions and speech. Some acoustical correlates. *Journal of the Acoustical Society of America* 52, 1238-1250.