

PROSODIC PARAMETERS FOR THE DETECTION OF REGIONAL VARIETIES IN ITALIAN

Elena Sardelli, Giovanna Marotta

University of Pisa (Italy)

elenasardelli@infinito.it, gmarotta@ling.unipi.it

ABSTRACT

The paper investigates the prosodic parameters which can be considered as relevant for the detection of regional varieties in spoken Italian. On the basis of acoustic analysis of three Italian varieties (Rome, Milan and Catanzaro), a crucial role is recognized in scaling and duration. Besides these features, tonal patterns and local pitch range also came out as being significant in variety discrimination through statistical distances. The same parameters are found relevant in sentence-type discrimination within the same variety. Results allow us to consider prosody as an important source of information for developing an automatic model in the geographical recognition of the speaker.

Keywords: Prosody – Variety detection – Scaling

1. INTRODUCTION

Different values of scaling in the same contexts have been traditionally considered as a secondary dimension of the speech signal. In [3], Bolinger already claimed that it might be related to the degree of the speaker's emotional involvement in the speech act. Within the autosegmental-metrical framework, the same parameter refers to the expression of different degrees in emphasis or prominence (e.g. see [2], [10]).

In Ladd [6] we find a detailed discussion about the relevance of pitch range variations, with special reference to downstep and raised peaks; the author suggests a prosodic "extrinsic but linguistic" category, involving "abstract relations between tones and between higher-level phonological constituents" [6: 278].

More recently, Prieto [10] has shown how tonal range variations can be significant in sentence-type information in Spanish: "H1 peaks of interrogative and imperative utterances are significantly higher than their corresponding statements, and exclamatory sentences have the highest peaks. Pitch-range variability is thus not exclusively used

in the Spanish language to convey differences in emphasis, but rather it shows categorical effects" [10: 4].

In previous studies regarding Italian (e.g. [8]), scaling was found as a relevant prosodic feature for geographical speaker recognition, inasmuch it was able to discriminate between Western Tuscan vs Central Tuscan varieties.

In this paper we would like to show the relevance of scaling in the detection of regional varieties in spoken Italian. At the same time, we will show how this parameter may play a crucial role in the expression of utterance typology within the same variety. Data are based on acoustic analysis and statistical results, which allow us to avoid subjective factors due to the speaker's perception.

Our final goal is the development and implementation of an automatic model for speaker recognition, exclusively referring to suprasegmental features, with possible applications in forensic phonetics, TTS and voice synthesis.

The empirical evidence is taken from Roman, Milanese and Catanzaro Italian, three varieties of Italian spoken in the Center, North and South of Italy, respectively.

2. SPEECH MATERIAL

In this study, 1768 files of spoken Italian have been analyzed. Speech material refers to *Map-Task* and *Difference Tests* dialogues of 6 male speakers per variety. Within this corpus, statements, continuative utterances, *yes-no questions* and *check-queries* have been selected.

The acoustic parameters analyzed are the following:

- a) quality of F0 patterns;
- b) scaling value on prominent, nuclear and final syllables;
- c) minimum and maximum F0 values, measured in the whole prosodic unit;
- d) relative duration of prominent, nuclear and final vowels.

Acoustic measurements have been run through software PRAAT after a manual labelling of prominent, nuclear and final syllables and vowels. Text grids have been annotated for each file.

3. DETECTION OF VARIETIES

Different scaling values as well as vowel durations can easily be recognized in all the utterance typologies we analyzed.

For the three varieties here considered, 497 *yes-no questions* have been processed. In this sentence-type, F0 final pattern is a fall associated with the last stressed syllable, followed by a rise on the final unstressed one. However, in spite of the same quality of movement (falling-rising), the following parameters can be considered as relevant for variety discrimination (see Table 1):

- the different scaling value associated with pitch accents (PA);
- the different scaling value associated with boundary tones (T%);
- the local pitch range (PR) value;
- the relative duration of involved vowels.

Table 1: Mean values of acoustic parameters analyzed in *yes-no questions*.

| | Rome | Milan | Catanzaro |
|-------------|---|---|---|
| Pattern |  |  |  |
| ST | 3,6 1,4 | 5,5 3,2 | 5,6 1,6* |
| Pitch Range | 6,63 | 8,84 | 9 |
| Min F0 | 107 | 93 | 110 |
| Fin F0 | 121 | 114 | 117 |
| *V duration | 112 | 109 | 138 |
| ToBI | H*+L H% | H+L* H% | (H+L)* |

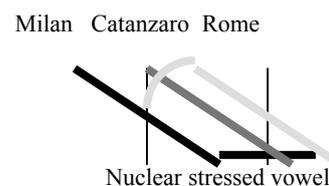
The Catanzaro variety can easily be recognized as the one often having a deletion or weakening of the unstressed final vowel. At the same time, the nuclear stressed vowel is significantly lengthened; for instance, compare vowel duration in Catanzaro (i.e. 138 ms) with that of Milan and Rome (109-112 ms, respectively).

As far as scaling is concerned, in Roman speech the melodic pattern is more compact, and located into higher levels of the melodic range, while in Milanese it reaches a wider span, which goes down towards the lowest level compatible with the speaker's pitch range. The respective mean values are 6,3 semitones (henceforth ST) for Roman speech and 8,8 ST for Milanese. According to our results, to detect different varieties in Italian, a tonal

excursion of approximately 2 ST is needed, whereas a wider span is necessary for utterance typology information (see below, § 4).

As shown in Table 1, not only did duration and scaling turn out to be significant for variety detection, but tonal alignment, too. The relevance of alignment in the discrimination of different accents has already been demonstrated by Atterer and Ladd [6] for German varieties. The PA notation reflects the different alignment in the three varieties of Italian: H*+L for Roman, (H+L)* for Catanzaro and H+L* for Milanese Italian (see Figure 1). The falling movement respectively starts before the stressed syllable left edge in Milan, at the beginning of it in Catanzaro and within it, in Rome. For the Catanzaro variety, we adopted a special transcription to indicate the perfect alignment of both tonal targets with the stressed syllable, as suggested by Marotta in [7].

Figure 1: Alignments of the PA H+L*, (H+L)* and H*+L, in *yes-no questions*.



Although the final boundary tone remains the same (i.e. H% for the three varieties), different scaling values come out as relevant for variety discrimination. In the Catanzaro speech, less than 20% of *yes-no questions* show the production of a final unstressed vowel (see the starred value in the first row of Table 1).

Different scaling values and different alignments between segments and tones are evident for variety detection in *check-queries* too. As shown in Table 2, F0 minimum value comes out as being a relevant parameter in the discrimination between the mean values for Roman and Milanese; for instance, compare Rome and Milan in Table 2 (126 Hz vs 97 Hz), then Rome and Milan in Table 1 (107 Hz vs 93 Hz). The Catanzaro data are not mentioned because this variety shows a different pattern in this particular kind of utterance.

Moreover, in *check-queries* of Milanese, we noticed a significant lengthening of final unstressed vowels, associated to two different tonal levels. On the spectrograms, a second articulation of the segment is clearly visible and it occurs exactly at the tonal turning point (beginning of a final F0

rising movement). Although this special alignment could be explained by resorting to the physical need of concentrating effort in dual pitch changes (see [4, 5]), we believe it is an interesting case of interference between segmental and suprasegmental features, with the effect of a pragmatic shifting (i.e. *yes-no questions vs check-queries*).

Table 2: Mean values of acoustic parameters analyzed in *check-queries*.

| | Rome | Milan |
|-------------|---|---|
| Pattern |  |  |
| ST | 6,1 6,7 | 3,7 1,8 |
| Pitch Range | 9,11 | 6,64 |
| Min F0 | 106 | 97 |
| Fin F0 | 139 | 115 |
| ToBI | H*+L H% | L+L* H% |

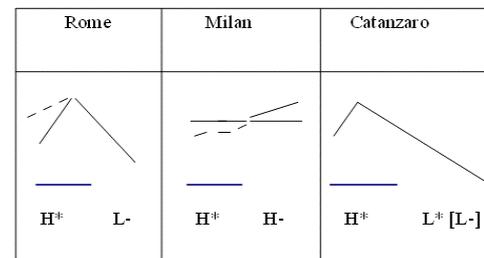
The same phenomenon occurs in some prominent vowels of statements in Catanzaro Italian, where vowel lengthening is associated with a wider pitch span and a true rearticulation of the segment.

In statements (856 files selected), our analysis confirms the results of previous research (see [6, 7, 9]), in recognizing scaling as a prominence marker. In all three varieties, prominence is associated with a high PA H*, followed by a phrase accent, which is highly variety-dependent (in both direction and scaling values: L-/H-).

As can be seen in Figure 2, the phrase accent is H- in Milanese Italian, whereas Roman and Catanzaro speakers perform a same phrase accent L-, although with a different scaling value. The mean value of the falling movement is about 7,5 ST in Rome and 11,1 ST in Catanzaro, where the pattern usually ends by the next stressed vowel. Utterance pitch range comes out as being variety-specific as well: 9,8 ST in Milan, 11,1 ST in Catanzaro and 11,3 ST in Roman Italian.

Furthermore, marks of the geographical origin of the speaker were found in the prosodic features of continuative utterances. In Milanese and Roman Italian – as occurs in many natural languages –, continuation is expressed by a rising final boundary tone H%, having a tonal range of 4,7 ST in Milan and 7,4 ST in Rome. In the Catanzaro speech, although the final boundary tone is usually L%, the final F0 is always higher than the mean value within the utterance.

Figure 2: Tonal patterns and PA associated with prominent syllables in statements.



In this Southern variety of Italian, continuation is marked by a particular nuclear PA, here transcribed as (L+L)*. Therefore, a range of 2,5 ST associated with the stressed syllable seems to be specific both for variety and utterance-type (see below, § 4).

As we have already stated (cf. § 1), our final goal is to develop an automatic model of recognition of the speaker's accent with exclusive reference to suprasegmental features. Having this special purpose, we decided not to run perceptual tests for relevant parameters detection (*versus* [9]), but to focus on statistical analysis, in order to identify, if possible, significant distances among varieties.

Tests have been run through all prosodic parameters we presented (scaling on selected contexts, mean values of minimum, maximum, initial and final F0, vowel duration), although only the duration and scaling came out as being statistically significant. For instance, in *yes-no questions* we found that a statistical distance greater than 1,6 is sufficient to discriminate between the three Italian varieties.

4. DETECTION OF UTTERANCE-TYPE

On the grounds of our results relative to the same variety of Italian (see Tables 1 and 2), we are able to assign relevance in sentence-type information to the parameters of scaling, local pitch range and relative minimum F0 value. For instance, within the Roman variety, we found a more compact pattern and higher values in *yes-no questions* with respect to *check-queries*.

Moreover, we noticed that alignment also has to be considered relevant: perceptual tests carried out through manipulation of F0 contour, show how a different alignment (e.g. H+L* and H*+L in the Roman variety) can definitely contribute to the shifting of the meaning from a question to a

statement. A similar effect is caused by duration and tonal span: as already mentioned in § 3, in *check-queries* we found the final unstressed vowel to be longer and often associated to two tonal levels. This particular status of the segment can shift a *yes-no question* into a *check-query*.

With reference to our data, a difference of 2 ST is needed for variety detection (see § 3), while a wider span is necessary to get relevant information about utterance-typology.

Our research clearly confirms what already has been attested in literature concerning questions (see [11, 13]), i.e. a higher initial level of F0, plus a sensitive lengthening of the nuclear stressed vowels, possibly caused by a wider span of the melodic pattern associated to them.

As regards to continuative utterances, in Roman Italian, the final rising movement always reaches a higher tonal level than the mean F0 value. In Milanese, on the other hand, the pattern is similar to that performed in the case of prominence (see Figure 2) and coordination, although the discriminative pattern comes out as being scaling once again. As a matter of fact, prominence is marked by a H* and a following plateau; a further rise up, even of a small range (1,5 ST) occurs in coordinative utterances, while continuation contours show a higher scaling value (4,7 ST).

In order to represent the different scaling values in relation to sentence-typology, we propose to use the diacritic [↑], already introduced by Walters in [14], i.e. ↑H%, for the final boundary tone in Milanese continuative utterances. However, in our proposal, the symbol doesn't refer to a specific value, as in [14], but rather it marks pragmatic information which is language-specific.

4. CONCLUSION

Our analysis has shown that scaling can be used as a prosodic cue for the discrimination among Italian varieties, but it may also function as a phonological parameter for utterance-type detection. From a mere phonetic point of view, we were able to confirm previous studies (e.g. [1]), regarding the relevance of the alignment in the identification of regional varieties. Furthermore, scaling, duration and local pitch range have been proved as being relevant suprasegmental parameters for sentence-type discrimination.

In order to represent some complex prosodic patterns, we introduced two new PA transcriptions: (H+L)* (for Catanzaro *yes-no questions*) and

(L+L)* (for Catanzaro continuative utterances), following what was proposed in [7]. Moreover, the diacritic [↑] has been used to mark a sensitive difference in scaling values between coordinative and continuative contours.

First analyses run to find statistical distances among varieties have shown that scaling and duration can be reliable prosodic parameters in the recognition of the geographical origin of the speaker. We aim at extending our analysis to other varieties of Italian, in order to develop an automatic model for speaker recognition based on suprasegmental features, as attempted in [12].

References

- [1] Atterer, M., Ladd, D. R. 2004. On the Phonetics and Phonology of "Segmental Anchoring" of F0: Evidence from German, *Journal of Phonetics*, 32, 177-197.
- [2] Beckman, M.E., Pierrehumbert, J.B. 1986. Intonational Structure in Japanese and English. *Phonology Yearbook* 3, 255-309.
- [3] Bolinger, D.L. 1982. *Intonation and its parts*. Stanford: Foris Publications.
- [4] House, D. 1990. *Tonal Perception in Speech*. Lund: University Press.
- [5] House, D. 2005. Phrase-final rises as a prosodic feature in wh-questions in Swedish human-machine dialogue. *Speech Communication* 46, 125-147.
- [6] Ladd, R. 1996. *Intonational Phonology*. Cambridge: CUP.
- [7] Marotta, G. 2000. Allineamento e trascrizione di toni accentuali complessi: una proposta. In: Locchi, D., *Atti delle X Giornate di Studio del GFS*, Napoli, 139-149
- [8] Marotta, G., Calamai, S., Sardelli, E., 2004. Non di sola lunghezza. La modulazione di F0 come indice socio-fonetico. In: De Dominicis A. et al. (eds.), *Atti delle XIV Giornate del GFS*, Roma, Esagrafica, 210-215.
- [9] Peters, J., Gilles, P., Auer, P., Selting, M. 2003. Identifying regional varieties by pitch information: A comparison of two approaches. *Proc. 15th ICPHS*. Barcelona, 1065-1068.
- [10] Pierrehumbert, J.B. 1980. *The phonology and phonetics of English intonation*. PhD thesis. MIT. IULC.
- [11] Prieto, P. 2003. Scaling of H1 peaks in Spanish: evidence from five sentence-types. *Proc. 15th ICPHS*. Barcelona.
- [12] Sardelli, E., Quali parametri prosodici per il riconoscimento automatico del parlante? In: *Atti della Conferenza Nazionale TAL*, Roma, 47-63.
- [13] Ultan, R. 1978. Some General Characteristics of Interrogative Systems. In: Greenberg, J. H., Ferguson, C. A., Moravcsik, E. A. (eds), *Universals of Human Language*, 4. 211-48. Stanford: Stanford University Press.
- [14] Walters, R. J. 2003. On the Intonation of a South Wales, Valleys Accent of English. *J. Int. Phon. Ass.* 33, 2. 211-238.