Prosodic variation and perceptive distinctness?
An experiment with some dialectal varieties of Italy

Valentina De Iacovo[1], *Paolo Mairano*[2]

[1]University of Turin, Italy
[2]University of Turin, Italy
v.deiacovo@gmail.com, paolo.mairano@unito.it

Abstract

Prosody represents a fundamental cue in the geographical identification of many Italian dialectal varieties. Especially in the interrogative modality, prosody can vary according to specific areas. Previous studies, based on the identification of specific prosodic functions, confirmed the fundamental role played by prosodic cues alone. But what happens when prosodic variations are not so obvious? Starting from five dialectal varieties spoken in Italy, we tested whether details characterising dialectal varieties are perceptively distinguishable. We devised an auditory discrimination task with delexicalized stimuli reproducing prosodic cues only. The results show that participants are able to discriminate dialects irrespective of the amount of prosodic distance.

Index Terms: dialectal varieties, prosodic cues, perception, dialectology, Italian dialects.

1. Introduction

1.1. Perceptive tests and dialectal identification

Prosodic variation represents an asset in many dialectal varieties. Studies [1] and [2] conducted on the regional varieties of Italy, showed, by means of different approaches, the existence of a great prosodic variability related to a specific dialectal area. For some varieties, nevertheless, pitch, duration and rhythm play a similar role, with the result that the prosodic distinction is less sharp.

Previous studies confirmed that prosodic cues are relevant in identifying questions and statements [3], different languages [4], regional varieties [5], but also that there is a certain margin within which changes can be made without any effect on perception [6].

Working on dialectal varieties spoken in Italy, our aim is to compare intonation patterns and show which prosodic features contribute to make a specific variety sound different from another one. Since instrumental analysis of acoustic features shows a small amount of variation for some varieties, our aim is to test if these changes are perceptively relevant or not.

2. Method

2.1. Stimuli

The test consisted in a discrimination task on the basis of auditory stimuli. For each trial, participants listened to a pair of synthesized stimuli. Stimuli were as delexicalized re-synthesis starting from 3 questions, each produced in 5 different dialectal varieties of Italy. For each trial, the participant had to choose if the two stimuli (i.e. re-synthesized questions) were asked by people speaking the same dialectal variety. As we already mentioned, among the varieties previously analysed, some questions presented a very small variation in the final part (corresponding to the verb and the object), so we wanted to test whether these differences were perceptively relevant. We therefore chose 5 dialectal varieties (Roma, Frosinone, Prato, Taranto and Pollina), where the first three (Prato, Roma and Frosinone) and the last two (Taranto and Pollina) presented a very similar pattern (see fig. 1). Also, we wanted to test if the position of the stress was relevant to identify a certain variety; so we chose three different neutral questions, always composed of a bisyllabic verb and an object. The object was composed by an article and a trisyllabic word, with three different stress pattern: oxytone (code *oxy*), paroxytone (code *par*), proparoxytone (code *pro*).
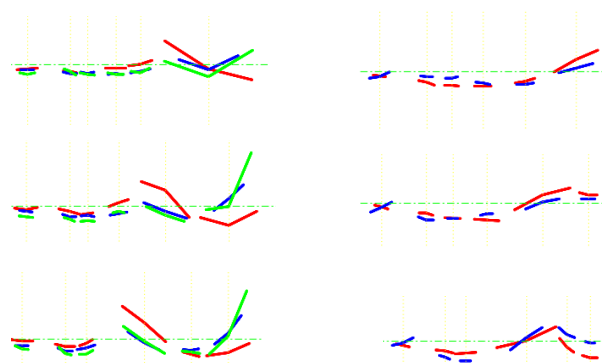

Figure 1: Intonation curves of the varieties of Roma, Frosinone and Prato (left column), Taranto and Pollina (righ column).The last trisyllabic word is oxytone (first row), paroxytone (second row) and proparoxytone (third row).

Since the original questions had been produced by 5 different speakers belonging to the 5 different varieties analyzed, we normalised only the pitch values of the 15 questions using a MatLab script in order to compensate differences in pitch mean across speakers. Samples were therefore delexicalised and re-synthesised to preserve the duration, the relative pitch changes and the intensity of the original WAV files: in short, they reproduced only the prosodic information. As we already mentioned, we only considered the last part of the sentences, corresponding to the verb and the object: this choice was made in the effort to minimize other potential differences such as thematisations/rhematisations; this allowed participants to concentrate exclusively on a precise part of the question, presenting the variation we wanted to test. In summary, the

test included 5 varieties (Roma, Frosinone, Prato, Taranto, Pollina, respectively coded *Rm, Fr, Pr, Ta, Po*) and, for each of these, 3 stress patterns, for a total of 15 possibilities.

2.2. Format of the test

For each trial, stimuli were matched according to four possible combinations:

1. same variety, same typology (15 combinations) → for example, Ta*par*Ta*par*;

2. same variety, different typology (30 combinations) → for example, Fr*pro*Fr*par*;

3. different variety, same typology (60 combinations) → for example, Rm*par*Pr*par*;

4. different variety, different typology (120 combinations) → for example, Ta*pro*Po*par*.

In order to have a balanced amount of stimuli, we doubled trials for cases 1 and 2, for a total of 90 stimuli. Conversely, we reduced the number of trials for case 4 by employing only the combinations with *par*/*pro*, plus 10 trials, for a total of 30 trials. Each trial consisted of two stimuli spaced out by a pause of 600 ms, for a total of 180 trials. Before the test, each participant performed a training session including 4 trials that were excluded from the analysis. The results of each participant were therefore stored in a csv file generated by the application at the end of each session.

2.3. Set up of the test

The test has been completed through an application conceived for Apple iPad. Before starting the test, the participants filled a short questionnaire with personal data (sex, age, education) and linguistic knowledge (their mother-tongue, whether they and their parents speak a dialect, how many foreign languages and which proficiency level they have. Most of the participants were students. They did not receive any monetary compensation for taking the test, but they received a partial credit for the exam in General Linguistics. The tests took place in the Laboratory of Experimental Phonetics "Arturo Genre" of Turin. Before starting the test, the first author personally explained how the test was conceived and they were invited to ask any questions after reading the instructions. The listeners sat in a testing sound-proof booth in front of an iPad Mini 4 equipped with Sennheiser HD451 headphones; after reading the instructions, they started a training session with 4 trials and then proceeded to the real test session.

On each trial, they firstly heard the two synthetic stimuli and then pressed a key to indicate if the stimuli heard belonged to the same (*yes*) or not (*no*). The interface used was extremely simple, consisting exclusively of the *yes* and *no* buttons (fig.) so that participants only had to press the button corresponding to their answer and they were immediately put forward to the next audio sample. The stimuli were presented randomly and the participants could only hear each stimulus once before making their response without any possibility to go back and correct the answer. The participants did not receive any feedback about their responses, but at the end of the test they could ask for their 'score' and could give us their impressions about the test. At that time, we provided more detail about the research that we carried out.

During the test, there were two break times (one every 60 stimuli) to give the participants the time to pause and continue the test when they wanted. At the end of the test, a new screen appeared and, finally, the application created a CSV file containing the stimulus order, the correctness of the answer, the code stimulus, the response given, the reaction time, the name of the test, the code to identify the accuracy of the answer, the number associated to the stimulus and the replay number.

2.4. Participants

The experiment has been submitted to 20 students (10 males, 10 females). All participants were native Italian speakers, aged between 23 and 30. The majority claimed to have no active competence of any dialectal variety, while all but two of them claimed to have passive competence of at least one dialectal variety of Italy.

3. Results

3.1 Statistical analysis

The data were analysed via a logistic mixed-effects model (responses) and a linear mixed-effects model (response time). The independent variables were accuracy of responses and response time. Both models included 2 fixed effects: dialect similarity (same dialect vs similar dialect vs different dialect) and stress pattern (same structure vs different structure). 'Same dialect' refers to trials where stimulus 1 and 2 are re-synthesised questions in the same dialect; 'similar dialect' refers to trials where stimuli 1 and 2 are re-synthesised questions in dialects with similar pitch contours' (e.g. Rome and Frosinone); 'different dialect' refers to trials in which stimuli 1 and 2 come from dialects with different pitch contours (e.g. Rome and Taranto). Similarly, 'same structure' refers to trials in which stimuli 1 and 2 had the same stress pattern on the last word, whereas 'different structure' refers to trials in which stimuli 1 and 2 had different stress patterns on the last word (e.g. oxytone vs paroxytone). The models also include 2 random effects, namely participant and stimulus pair. The effects of the independent variables were estimated with ANOVA table of type III with Satterthwaite approximation for degress of freedom, as provided by the lmerTest package in R [12].

3.2 Accuracy of responses

Responses given by participants are plotted in fig. 2 by dialect similarity (x axis) and stress pattern (colour). We tested the statistical significance of these two variables via a logistic mixed-effects model with dialect similarity and stress pattern as fixed effect, participant and stimuli as random effects. The output tells us that the main effects of both dialect similarity and stress pattern are highly significant ($p < .001$ for both). However, post-hoc comparisons with Bonferroni correction reveal that statistical differences exist only between responses for same dialect vs different dialect ($p < .001$) and same dialect vs similar dialect $p < .001$), but not between responses for different vs similar dialect ($p = .52$). This means that, on the one hand, trials in which the two stimuli were re-syntheses of the same dialect were recognized as such, with higher ratings than trials containing similar or different dialects. On the other hand, no significance difference occurred between trials containing dialects with similar vs different pitch

contours, thereby suggesting that small prosodic details are sufficient to trigger perceptual distinction.
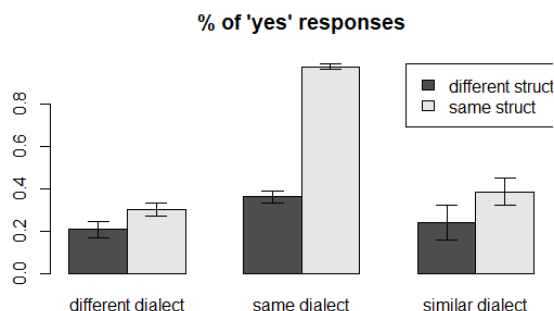
% of 'yes' responses

Figure 2: *Participant responses by dialect similarity and stress pattern*

3.3 Response time

Response time are plotted in fig. 3 by dialect similarity (x axis) and stress pattern (colour). Outliers beyond the .05 and .95 percentiles have been discarded from the charts and from the analysis.
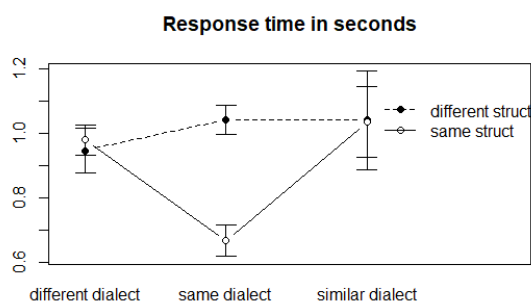
Response time in seconds

Figure 3: *Response times by dialect similarity and stress pattern*

We tested the statistical significance of these two variables via a linear mixed-effects model with dialect similarity and stress pattern as fixed effect, participant and stimuli as random effects. We found significant main effects for dialect similarity ($p < .001$) and stress pattern ($p < .001$). However, post-hoc pairwise comparisons with Bonferroni correction show that the only condition responsible for statistical significance is 'same dialect and same stress pattern' ($p < .001$). In other words, response times for this condition are significantly distinct from response times for all other conditions; but all other conditions do not differ significantly from one another (all $p = .27$ or higher). This is also visible in the chart above, where the marker for 'same dialect' and 'same structure' clearly indicates lower response times than the rest. This suggests that participants found it easier to recognize re-synthesized stimuli from the same dialect and with the same stress pattern than any other pair of stimuli. Again, no statistical difference was found between stimuli belonging to dialects with similar vs different pitch contours, although response times do seem to be descriptively higher for the latter. This leads us to hypothesize that dialects with similar

pitch contours were potentially more problematic to distinguish, and this caused slightly longer response times, although our current data do not reach statistical significance.

4. Discussion and conclusion

Results have shown the main effects of dialect similarity and stress pattern. When the dialectal variety and the stress pattern are both the same, there is a maximum percentage of accuracy; conversely, when the dialectal variety is the same but the stress pattern changes, accuracy decreases to less than 40%. Results are comparable when two similar dialectal varieties are presented with the same structure: the ability to distinguish two similar varieties is lower and this is also reflected by the response time, although results do not reach statistical significance in this case. The most relevant finding is the fact that the discrimination of dialects presenting similar pitch contours does not seem to differ from the discrimination of dialects presenting very distinct pitch contours. This suggest that small prosodic details can be sufficient to perceptive discrimination of dialectal variants.

This experiment represents only a starting point for the investigation of the prosodic perception of dialectal varieties. Further tests will be conducted in the future, considering a greater amount of varieties and a more consistent group of participants.

5. Acknowledgements

6. References

[1] L. Canepari, "L'intonazione. Linguistica e paralinguistica.", Napoli: Liguori, 1985.

[2] R. Endo, & P.M. Bertinetto, "Aspetti dell'intonazione in alcune varietà di italiano", *Fonetica e fonologia degli stili dell'italiano parlato. Atti delle VII giornate di studio del Gruppo di fonetica sperimentale (AIA)*, pp. 27-49, 1997.

[3] F. Ferrero, "Studio preliminare sulla classificazione percettiva di contorni intonativi sintetici come interrogativi/non interrogativi", *Atti del Convegno della Società di Ling. Italiana*, pp. 81-93, 1976.

[4] J.J. Ohala, & J.B. Gilbert, "Listeners' ability to identify languages by their prosody", *Problèmes de prosodie, Vol. II: Experimentations, modèles et fonctions*, pp.123-131, 1981.

[5] F. Schaeffler, & R. Summers, "Recognizing German Dialects by prosodic features alone," *Proceedings of the 14th ICPhS*, pp. 2311-2314, 1999.

[6] E. Garding, & A.S. Abramson, "A study of the perception of some American English intonation contours", *Studia Linguistica*, vol.19, 1965.

[7] M. Contini, J.-P. Lai, A. Romano, S. Roullet, L. de Castro Moutinho, R.L. Coimbra, U. Pereira Bendiha, & S. Secca Ruivo, "Un projet d'atlas multimédia prosodique de l'espace roman", *Proceedings of the International Conference Speech Prosody 2002*, pp. 227-230, 2002.

[8] L. Canepari, "Italiano standard e pronuncia regionale.", Padova, CLEUP, 1983.

[9] P. Lieberman, "On the Acoustic Basis of the Perception of Intonation by Linguists", *WORD*, 21:1, 40-54, 1965.

[10] W. Majewski, & R. Blasdell, "Influence of fundamental frequency cues on the perception of some synthetic intonation

contours.", *The Journal of the Acoustical Society of America*, vol. 45, no. 2, pp. 450-457, 1969.

[11] P. Delattre, "Les dix intonations de base du français", *The french review*, vol. 40, no. 3, pp. 1-14, 1966.

[12] A. Kuznetsova, P.B. Brockhoff, & R.H.B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models", *Journal of Statistical Software*, 82 (13), 1-26, 2017.