

LOW TONE VERSUS 'SAG' IN BARI ITALIAN INTONATION; A PERCEPTUAL EXPERIMENT

Martine Grice¹ and Michelina Savino^{1,2}

¹Institut fuer Phonetik, Universitaet des Saarlandes, Germany

²Istituto di Filosofia e Scienze del Linguaggio, Universita di Bari, Italy

ABSTRACT

A perceptual test shows that the presence of a f0 dip just before the accented syllable plays a role in the perception of a yes-no question as opposed to a command in Bari Italian, and that this dip may be as small as 20 Hz. It is represented as L in a L+H* pitch accent.

INTRODUCTION

In Italian, both in the standard and in many regional varieties, including that of Bari, there are no syntactic or morphological means of distinguishing an information-seeking yes-no question from a non-question, such as a statement or command; instead, they are distinguished by means of intonation. For example, the sentence "Lo mandi a Massimiliano" can be interpreted as a question, loosely translated as "Will you send it to Massimiliano?", or a command "Send it to Massimiliano!" or statement "You send it to Massimiliano", depending on its intonation pattern. The actual pattern used differs from one local variety to another. Analysis of dialogue recordings involving six speakers of Bari Italian (BI) within a Map Task framework [1] has indicated that information-seeking questions (queries) have a rising-falling nuclear contour (see fig.1). Rising-falling contours also occur in questions of this type in Palermo Italian [2], [3], and in a number of other regional varieties of Italian [3]. We analyse the rising-falling pattern as a rising pitch accent on the focussed item, followed by a low target at the phrase boundary. The peak is around the middle of the accented syllable, and the preceding valley is on the preceding syllable. BI commands (instruct moves in [1]) have a falling nuclear contour where the fall begins

around the beginning of the accented syllable (see fig.2). Within an autosegmental approach, we analyse the query contour as L+H* L-L% and the command contour as H* L-L%.

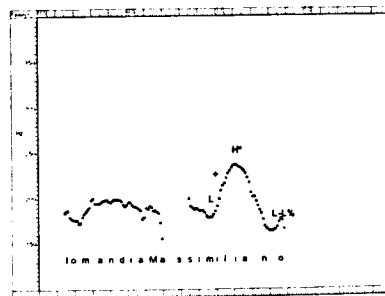


Figure 1. Natural rendition of query "Lo mandi a Massimiliano?"

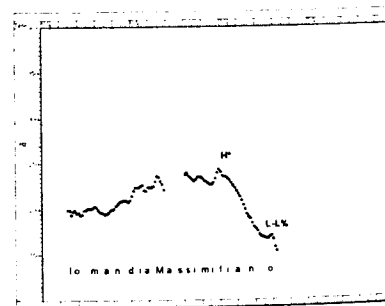


Figure 2. Natural rendition of command "Lo mandi a Massimiliano!"

Other autosegmental studies of Italian intonation, [2], [4], also use a L+H* pitch accent, but not in questions. Palermo Italian has L+H* L-L% in non-final items in contrast to L*+H L-L% in yes-no questions [2]; Standard Italian has L+H* in certain focussed items in non-questions [4]. Pitch accents with unstarred leading tones (i.e. T+T*) are more

controversial in accounts of English (see [5] for discussion) than Italian. Not all autosegmental models of English intonation recognise L+H*; Ladd, for instance, treats L+H* as simply an emphatic version of H* [6]. One problem with the H* / L+H* distinction is that a H* pitch accent can be preceded by a degree of dip which is not attributed to a L tone, but rather to sagging interpolation. For example, between two H* tones which are far enough apart, Pierrehumbert's [7] model predicted sagging interpolation. This meant that a small dip would not be interpreted as a target L tone. Rather, it would be the result of an interpolation rule which applied automatically. A synthesis model [8] building on the foundation of [7] gave H* tones an underlying shape which, when filtered, created a peak accent with a small rise up to the local maximum. A L+H* pitch accent has a larger rise starting at a lower point.

Since it is clear that in B.I. the distinction between H* and L+H* is an important one, the aim of the perceptual experiment described in this paper is to investigate how low the local minimum has to be for a L tone to be perceived. The choice of sentence material precludes an analysis of the L tone as anything other than a leading unstarred tone of the nuclear pitch accent, since there is no potential prosodic boundary before the accented syllable 'LIA' of 'massimILIAno', so the minimum cannot be attributed to a boundary tone of any kind. Perceptual experiments have been carried out on f0 peaks for German [9] and Hungarian [10]. Below it is the height of a f0 valley or dip rather than a peak which is under investigation.

THE EXPERIMENT

Stimuli

A number of renditions of the sentence "lo mandi a Massimiliano" were produced by a female Bari Italian native speaker, both as a query and as a command. These tokens were analysed,

and a matching pair was selected where the tokens were closest in the f0 value of their endpoint and nuclear peak. F0 traces of the two tokens chosen are shown in figures 1 and 2. The f0 of each token was stylised using straight lines; and from the stylised tokens, two sets of stimuli were created using PSOLA resynthesis, as follows: Series Q: from the stylised query, 6 resynthesised versions were created as test stimuli by increasing (on a linear scale) the F0 value of the low target before the rise in 10 Hz steps up to the interpolation line between the two peaks; series C: from the original command, 6 resynthesised versions were created by lowering the F0 value at the same point in the preaccental syllable corresponding to the low target in the original query in 10 Hz steps down to a position close (within 6 Hz) to the original query. A small difference in height of the two nuclear peaks (3Hz) and a difference in height and position of the first peak of the two natural stimuli (see figures 1 and 2) meant that the degree of dip from the interpolation line could not correspond to exactly the same Hz value at each step in the two series. The total set of test stimuli was 16 (2 original utterances + 2 series x 7 stimuli per series). The way the stimuli were constructed is shown in figure 3.

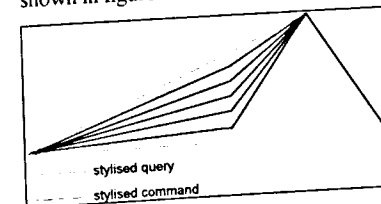


Figure 3. Schematisation of resynthesised continua.

Subjects

Fourteen native Bari Italian speakers took part in the experiment. They were between 20 and 40 years old and were students and staff at the University and Politecnico of Bari. None of them had a background in phonetics.

Experimental procedure

The difference between the two communicative functions, query and command, was explained by means of examples in context. The 16 stimuli were presented five times, each in a randomised sequence in blocks of ten. Each stimulus was preceded by a 250 ms warning tone and 1 second silence. 5 seconds of silence followed for the subjects to respond. After each block of 10 stimuli there was a larger 11 second pause and a double tone of 250 ms as a precursor to the next block of stimuli. After each stimulus, subjects indicated on an answer sheet whether the utterance they heard was a query or a command. The total test duration was circa 20 minutes.

RESULTS AND DISCUSSION

Figure 4 shows the percentage of "query" responses for the series of stimuli with the stylised version of the query as the base stimulus (series Q) as a function of the size of the dip in f_0 on the pre-accentual syllable 'mi'. Figure 5 shows the percentage of "command" responses for the series originating from the stylised command (series C).

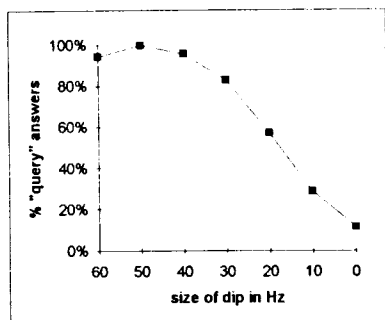


Figure 4. Percent "query" responses as a function of the size of the dip in Hz on the pre-accentual syllable. Base stimulus originally query (series Q).

As it has to be expected, in set Q there is a shift in response from query to command as the dip reduces in size, and

in series C there is shift from command to query as the dip increases.

In series Q, the level at which more than 50% of responses were "query" was at a dip value of between 10 and 20 Hz (57% "query" responses for 20 Hz dip).

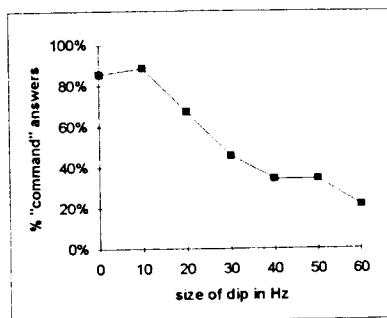


Figure 5. Percent "command" responses as a function of the size of the dip in Hz on the pre-accentual syllable. Base stimulus originally command (series C).

This means that there only has to be a very small dip (considerably smaller than that observed in many cases of typically sagging interpolation, as exemplified in [7]) for a query and therefore L+H* to be perceived. However, in creating the stimuli from natural renditions, we neither manipulated nor held constant all the parameters which might serve to make the distinction between query and command. It could be, therefore, that there were other cues in series Q which led to the perception of a query, other than simply the degree of dip. It was with a view to controlling for some of these effects that a command was taken as the basis for the second series (series C). In this series, the degree of dip at which more than 50% of responses were "query" lay between 20 and 30 Hz (46% command, i.e. 54% query responses for 30 Hz dip). This means that the dip has to be lower in stimuli where the base token was a command. However, a 30 Hz dip is still comparable with cases in [7] of sagging interpolation in a similar pitch range.

It is apparent that the response scores in series C are less extreme than those in series Q. The base token in series C is recognised as a command in 86% of cases, whereas the base token for series Q is recognised as a query in 94%. The command which has been maximally altered (60 Hz dip) is only recognised as a query 79% of the time whereas the query with no dip at all is recognised as a command 89% of the time. Two factors may have influenced this.

1. Commands often display a larger downtrend in their baseline values than the one selected here for manipulation (other commands with greater downtrends had to be rejected because their endpoints were too low). The downtrend may be due to final lowering, as discussed in [11] in relation to Japanese, or to declination, as discussed in [12] in relation to Danish. In both Japanese and Danish, the downtrend does not occur in questions but does in other utterance types. The apparent flatter baseline could have made the stimuli in series C sound less command-like.

2. The presence or absence of a dip in the natural f_0 contour appears to be linked to the position of the peak within the following accented syllable. After a dip, the peak is later (around the middle of the syllable) whereas without a dip, it is at the start of the vowel. Since the original peak position was retained in both series, it could have played a role in making the stimuli in series C sound less query-like, owing to the inevitably steeper rise from the dip. The cue for L+H* might not only be the presence of a dip in F_0 before the accented syllable but also a later peak.

We can only speculate at this stage that such factors as final lowering (or declination) and peak and dip position may have made series C responses less differentiated. Further experimentation is needed to investigate these parameters.

Since we took two base tokens, one query and one command, and constructed two series of stimuli, we have

confirmed that subjects can use a dip in f_0 as at least one cue for discriminating between questions and commands. We have also provided support for the hypothesis that the dip in f_0 constitutes a target L tone, rather than an automatically sagging interpolation between two peaks, and shown that the dip does not have to be very low to be perceived as a L tone. This has implications for theories of intonation which allow for non-monotonic interpolation.

ACKNOWLEDGEMENT

We would like to thank Bill Barry and Mario Refice for advice and comments on an earlier draft of this paper.

REFERENCES

- [1]Grice M., R. Benzmueller, M. Savino, B. Andreeva (this proc.), The intonation of queries and checks across languages: Data from Map Task dialogues.
- [2]Grice M. (1992), *The intonation of interrogation in Palermo Italian: implications for intonation theory*, PhD dissertation, University College London, also (1995) Niemeyer L.A. series 334.
- [3]Canepari L. (1992), *Manuale di pronuncia italiana*, Bologna: Zanichelli.
- [4]Avesani C. (1990), A contribution to the synthesis of Italian intonation, *ICSLP '90 Proceedings*.
- [5]Grice M. (in press), Leading tones and downstep in English, *Phonology* 12.2.
- [6]Ladd D. R. (1983), Phonological features of intonational peaks, *Language*.
- [7]Pierrehumbert J. (1980), *The phonology and phonetics of English intonation*, PhD dissertation, MIT.
- [8]Anderson et al. (1984), Synthesis by rule of English intonation patterns, *Proc. IEEE*
- [9]Kohler K. (1987), Categorical pitch perception, *Proc. XI ICPhS*, Tallinn.
- [10]Gosy M. and J. Terken (1994), Question marking in Hungarian: timing and height of pitch peaks, *J.Phon.* 22.
- [11]Pierrehumbert J. and M. Beckman (1988), *Japanese Tone Structure*, Cambridge: MIT Press.
- [12]Thorsen N. (1983), Two issues in the prosody of Standard Danish, in Cutler A. and D.R. Ladd (Eds.), *Prosody: Models and Measurements*, Berlin: Springer.