PERCEPTION OF QUESTIONS AND STATEMENTS IN NEAPOLITAN ITALIAN

Mariapaola D'Imperio and David House***

*Department of Linguistics, Ohio State University, 222 Oxley Hall, 1712 Neil Avenue
Columbus, OH 43210-1298, USA. e-mail: dimperio@ling.ohio-state.edu
**Department of Linguistics and Phonetics, Lund University, Helgonabacken 12, S-223 62 Lund, Sweden
e-mail: david.house@ling.lund.se

ABSTRACT

This paper addresses the problem of the perception of two different pitch accents in Italian which signal two utterance types (interrogative and declarative). The questions asked concern whether the major perceptual cue to this category distinction involves only the temporal alignment of the high level target with the syllable or if the category percept also depends on the presence of a rising or falling melodic movement within the syllable nucleus. The results show that the primary perceptual cue for questions is a rise through the vowel, while the primary cue for statements is a fall through the vowel. The results bear upon a general theory of intonation and our understanding of intonation in Italian as well as on current models of tonal perception in speech.

1. INTRODUCTION

In Italian, as in English, lexically stressed syllables can receive a pitch accent with the last accented syllable bearing the most prominent accent (i.e. "nuclear accent"). Different pitch accents can cue different pragmatic and grammatical functions. In both the standard and in the Neapolitan variety of Italian, unmarked declarative sentences are characterized by a falling nuclear accent which has been analyzed [1,2] as a sequence of a high target followed by a low one on the stressed syllable (H+L*). Results from a previous study [2] suggested that the accent peak is later in statements marked by narrow focus; in this case the nuclear accent can be labeled as H*. The shape of this accent is a local rise on the nuclear syllable followed by a fall within the vowel.

Differing from the standard variety [1], the shape of the canonical nuclear accent in Neapolitan Italian yes/no questions is a local rise on the nuclear syllable (L+H*), followed by a low tone marking the end of the phrase [3,4]. Therefore, as in the narrow focus statements, the nuclear accent of Neapolitan yes/no questions is characterized by a high tonal target, but this appears to be reached later in the stressed vowel than in statements. A similar pattern has been described for Palermo Italian [5] and Bari Italian [6], even though it appears that the high target is reached later in the syllable in Neapolitan Italian than in other previously described varieties. The timing characteristics of the target high tone appear then to be essential in discriminating yes/no questions from statements.

Recent work on the perception of melodic contours in speech [7,8] has questioned the assumption that melodic contours are exclusively encoded in terms of targets. It appears, instead, that in regions of high spectral stability tonal movement can be perceptually encoded as a contour (rise or fall). The present study tested two possible hypotheses about what determines the percept of a statement as opposed to a question. According to the first hypothesis, the latency of the high target (peak) relative to the onset of the accented syllable is the main cue for a category shift. This hypothesis, therefore, predicts that the later the peak is realized, the more question judgements will be produced.

However, it is possible that the difference between a H* and a L+H* would not simply concern the temporal alignment of the high level target with the syllable but would also heavily depend on the presence of a melodic rise for questions and a fall for statements. Therefore, the second hypothesis predicts that a clearly perceptible rise within the syllable nucleus is more important than the late timing of the high target as a cue for the identification of questions.

2. METHOD

Two tokens of the sentence "Mamma andava a ballare da Lalla" (Mom used to go dancing at Lalla's) produced with intended narrow focus on the last word by a native female speaker of Neapolitan Italian were selected from a corpus of read speech (cf. [2]). One of the tokens was produced as a statement and the other as a yes/no question. The sentences were judged as unambiguous representatives of their respective categories and differ in the timing of the final peak. In the declarative, the peak occurs early in the accented vowel, while in the question, the peak occurs at the end of the vowel (Fig. 1).

Fundamental frequency differences between the two tokens were neutralized by stylizing the F0 of each token using mean F0 values for the two utterances (Fig. 1). From these two base stimuli, twenty test stimuli were created using PSOLA synthesis. From the declarative base stimulus, five stimuli were created by shifting the

peak in four steps of equal duration (33 ms) forward through the vowel (*decl-peak*, Fig. 2A) and five stimuli were created by shifting the fall in four steps of equal duration (33 ms) forward through the vowel (*decl-fall*, Fig. 2B). From the interrogative base stimulus, five stimuli were created by shifting the peak in four steps of equal duration (35 ms) backwards through the vowel (*inter-peak*, Fig. 2C) and five stimuli were created by shifting the rise in four steps of equal duration (35 ms) backwards through the vowel (*inter-rise*, Fig. 2D).

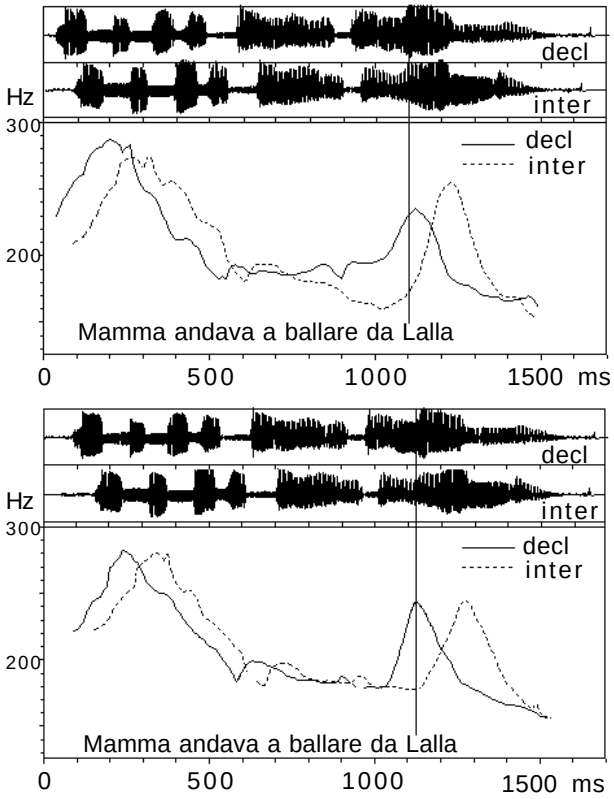

Figure 1. Waveforms and fundamental frequency contours of the two original utterances (upper panel) and of the utterances after tonal normalization (lower panel). The vertical lines indicate the temporal line-up point at vowel onset in "Lalla".

A test tape was created in which each of the twenty stimuli was presented four times. The stimuli were randomized in sequences of twenty and presented in blocks of ten, for a total of eight blocks. Each block was announced by its number on the tape with a five-second silent interval between the announcement and the first stimulus of each block. A silent interval of five seconds also followed each stimulus (response interval), and a ten-second interval separated each block from the announcement for the following block. Twenty native speakers of Neapolitan Italian, who were between 23 and 34 years old, participated in the experiment. Subjects were given written instructions and asked to indicate after each stimulus if it sounded most like a statement or a question by marking the appropriate box on an answer sheet. The test lasted approximately 15 minutes.

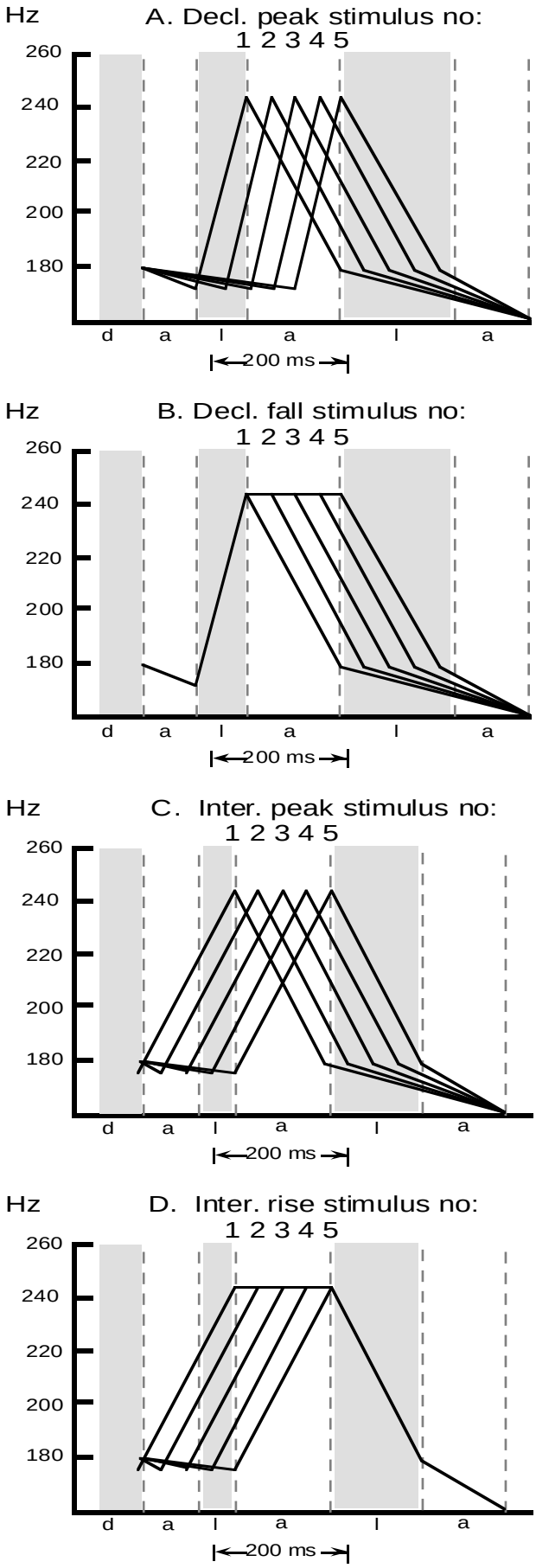

Figure 2. Schematized stimuli used in the perception test.

3. RESULTS

Figure 3 shows percent of "statement" responses across stimulus number for each series. As we can observe, responses to both the decl-peak and inter-peak stimulus sets present a sharp category boundary at the middle stimulus. Responses to the decl-fall series were largely declarative becoming ambiguous as the fall is moved late in the vowel. Responses to the inter-rise series showed a majority of interrogative responses and an increase in the number of interrogative responses as the rise became more prominent through the vowel.

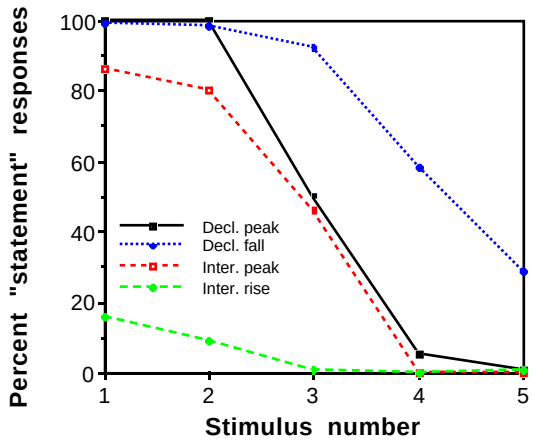

Figure 3. Results of the perception test plotted as percent statement responses for all stimuli.

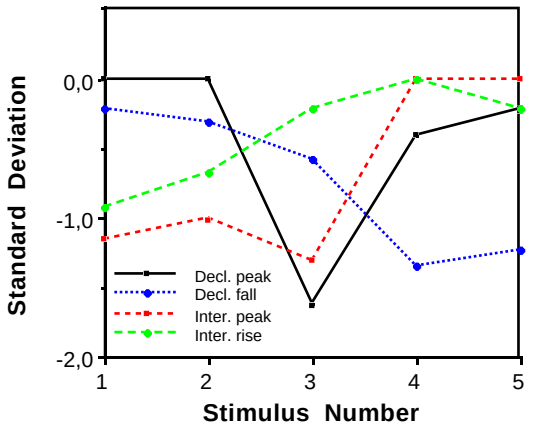

Figure 4. Standard deviations of the responses.

Figure 4 plots the standard deviations of the responses for each stimulus. The most consistent statement responses were elicited by the decl-peak 1 and 2 and the decl-fall 1 and 2, i.e. stimuli where there is a clear fall through the vowel. The most consistent question responses were elicited by the decl-peak 4 and 5, the inter-peak 4 and 5 and the inter-rise 3, 4 and 5, i.e. stimuli where there is a clear rise through the vowel.

4. DISCUSSION

The results clearly reveal that it is possible to switch the perceived category between questions and statements by moving the accent peak within the accented vowel, thereby replacing a fall by a rise or a rise by a fall in the accented vowel. Eliminating the rise or the fall from the vowel by shifting it into the neighboring consonant did not result in a category shift, but rather increased ambiguity. The results appear therefore to confirm the second hypothesis in that perceptually a rise in the vowel is the most important cue for the question and a fall in the vowel is the most important cue for the statement. The late fall in the syllable coda for the question and the early rise in the syllable onset for the statement are therefore seen as secondary cues as are peak height and rise gradient.

Since the duration and the timing of the rise in the inter-peak and the inter-rise stimuli of like number was the same, how can we explain their different outcomes in terms of the results? In the early inter-peak stimuli the fall in the vowel immediately following the rise conflicts with it as a cue to the pragmatic role of the utterance. A fall through the vowel is therefore a strong signal of a declarative utterance. The inter-rise stimuli, instead, contain a fall whose beginning and end are always realized within the boundaries of the syllabic coda, which makes the fall less clearly perceptible and eliminates it as a declarative marker. This might explain why, overall, the inter-rise stimuli elicited an overwhelming number of question responses even in those cases where the rise may not have been optimally perceptible because of intervening spectral discontinuities [7,8]. In a way similar to the inter-rise stimuli, the decl-fall stimuli are characterized by a rise that is not optimally perceptible, since it is produced within the consonantal onset of the accented syllable.

Even though the timing of the peak is exactly the same for equal stimulus number in the inter-peak and in the decl-peak series, we notice a bias for inter-peak to receive more question responses (the area underneath the curve in Figure 3 is smaller for inter-peak stimuli). This may in part be due to the slope of the rise, which is steeper in decl-peak stimuli, and to the duration of spectral instability within the signal for the two stimuli series. As we can see from Figure 2, the duration of the consonantal onset in the decl-peak stimuli is greater than in the inter-peak stimuli. It is possible that for inter-peak stimuli the region of spectral instability is not big enough to seriously compromise the percept of a rise.

It was not surprising that subject responses were overall less consistent in the inter-peak results than in the decl-peak results. Inter-peak stimuli are characterized by a peak that is surrounded by a rise and a fall of nearly equal slope and size. The simultaneous presence of a perceptible rise and fall produced highly conflicting cues, which can explain response inconsistency in the first three stimuli of the inter-peak series. The decl-peak stimuli, instead, are characterized by a rise that is always

faster and steeper than the subsequent fall. Hence, the rise does not conflict with the fall in the early decl-peak stimuli.

The perceptual data presented here offer additional and independent evidence for the relevance of the falling tonal movement in the perception of the category "statement with narrow focus". This led us to include L within the pitch accent label for this category, in a way similar to Grice [5]. Grice argued that the L tone is indeed an integral part of the pitch accent on the basis of the fact that the endpoint of the fall is always consistently aligned relative to the H; this observation is also true for Neapolitan Italian (see D'Imperio [2, 4]). Therefore we propose a unified analysis of statement pitch accents as consisting of an HL sequence independent of breadth of focus, where H+L* will be used as a label for broad focus utterances and H*+L, instead of monotonal H*, would then be employed for labeling narrow focus ones.

Perception data for yes/no questions suggest that no category difference exists due to breadth of focus [9]. Therefore we propose a unique analysis of the question pitch accent as L+H*. A further motivation for attributing the association of H to the stressed syllable is that, in production, L does not have to be realized within the boundaries of the stressed syllable, but is often realized within the preceding syllable. Moreover, L is not sustained throughout the stressed vowel; instead, the rise starts soon after the onset of the stressed syllable, so that H will be reached around the offset of the vowel in the prototypical case. This analysis appears to be supported by perceptual data presented here. In order for the question category to be identified, the H does not necessarily have to be realized in the later part of the stressed vowel. If we look at the data for the rise-shift stimuli, we notice that stimulus number 3, in which the H is reached in the middle of the vowel, elicited nearly 100% question responses (see Fig. 3).

We certainly cannot totally exclude the existence of L*+H in Neapolitan Italian; however, no meaning contrast between L+H* and L*+H can be imagined at this point. Therefore, we believe that the LH sequence is what matters for questions to be identified, independent of the exact timing of the H, which renders the choice of label relatively irrelevant.

5. CONCLUSION

The results of this perception experiment indicate that the primary cue for questions in Neapolitan Italian is a rise through the accented vowel, while the primary cue for statements is a fall through the accented vowel. These results have implications for our general understanding of the perception of melodic movement and pitch accents and add evidence in support of the perceptual importance of pitch movement through areas of spectral stability for identifying tonal categories. The results also have implications for the labeling of tones in Neapolitan Italian and are seen as an important first step toward a more extensive perceptual study of pitch accent categories in Neapolitan Italian.

6. ACKNOWLEDGMENTS

We are very grateful to all the members of CIRASS labs, University "Federico II" of Naples, for making the experiments possible; to all of our subjects, and most especially to Prof. Federico Albano Leoni and Dr. Franco Cutugno. Special thanks to Marcus Filipsson and Brigitta Lastow at Lund University for help with the synthesis programming and for assistance in processing the results.

7. REFERENCES

[1] Avesani, C. "A contribution to the synthesis of Italian intonation", Proc. *ICSLP 90, vol. 1*:833-836, Kobe, Japan, 1990.

[2] D'Imperio, M. "Timing differences between prenuclear and nuclear pitch accents in Italian", Poster presented at the 130th Meeting of the Acoustical Society of America, 27 Nov. - 1 Dec. 1995, St. Louis, MO, USA (*JASA, vol. 98, No. 5, Pt. 2,* p. 2894), 1995.

[3] Caputo, M.R. "L'intonazione delle domande si/no in un campione di italiano parlato", *Atti delle 4e Giornate di Studio del GFS (A.I.A.)*, pp. 9-18, Torino, Italy, 11-12 novembre, 1993.

[4] D'Imperio, M. "Caratteristiche di timing degli accenti nucleari in parlato italiano letto", *Atti del XXIV Convegno Nazionale dell'Associazione Italiana di Acustica*, pp. 55-60, Trento, Italy, 12-14 giugno, 1996.

[5] Grice, M. *The intonation of interrogation in Palermo Italian: implication for intonation theory* (Linguistische Arbeiten Nr. 334). Tübingen, Niemeyer, 1995.

[6] Grice, M., Benzmüller, R., Savino, M., Andreeva, B. "The intonation of queries and checks across languages: Data from map task dialogues", Proc. *ICPhS 95, vol. 3*:648-651, Stockholm, Sweden, 1995.

[7] House, D. *Tonal Perception in Speech*, Lund: Lund University Press, 1990.

[8] House, D. "Differential perception of tonal contours through the syllable", Proc. *ICSLP 96, Fourth International Conference of Spoken Language Processing*, pp. 2048-2051, Philadelphia, PA, USA, 1996.

[9] D'Imperio, M. "Breadth of focus, modality and prominence perception in Neapolitan Italian", to appear in K. Ainsworth-Darnell and M. D'Imperio (eds), *OSU Working Papers in Linguistics 50*, 1997.