

Methods and Tools for Prosodic Analysis of a Spoken Italian Corpus

Michelina Savino*, Mario Refice*, Domenico Daleno⁺

* Dipartimento di Elettrotecnica ed Elettronica, Politecnico di Bari
⁺Dipartimento di Informatica, Università di Bari
via Orabona 4 – 70125 Bari, Italy
{esavino, refice}@poliba.it, daleno@di.uniba.it

Abstract

In the last few years, a number of actions has been carried out in Italy with the goal of collecting, annotating and making available a considerable amount of data of spoken Italian varieties. After a first phase, in which the AVIP corpus has been collected and transcribed at both segmental and suprasegmental levels, now research efforts have been concentrating on corpus analysis, starting from two preliminary yet crucial aspects, namely: a) developing strategies and software tools for controlling the semantic coherence of the AVIP database; and b) designing a DBMS scheme for allowing easy access to the data and for rendering the results of the online queries in a user-friendly manner, also by means of special graphical interfaces. In this paper both aspects are presented and discussed, focussing on the prosodic analysis of the database, in terms of the methodologies followed in the intonation labelling phase as well as the consequent strategies adopted in the implementation of software tools for prosodic analysis.

1. Introduction

In the last few years, a number of actions at a national level has been carried out having the goal of gathering, annotating and making available a considerable amount of data of spoken Italian varieties (Albano Leoni et al., 1998; Refice et al., 2000). During the first phase, the AVIP (Archivio di Varietà di Italiano Parlato, Spoken Italian Varieties Archive) corpus of spontaneous Map Task dialogues (Anderson et al., 1991) for Bari, Naples and Pisa varieties of Italian have been collected, orthographically transcribed, segmented and hand-labelled at both segmental and suprasegmental levels, and partially annotated (starting from the orthographic transcriptions, i.e. without temporal alignment to the speech signal) at the pragmatic/textual level. Suitable software tools for assisting human transcribers in labelling procedures and for building the database have been developed within the project (Refice et al., 2000).

In the running second phase, research efforts have been concentrated on corpus analysis by taking into account two important preliminary aspects, namely:

- developing strategies and software tools for controlling the formal and semantic coherence of the database;
- designing a Data-Base Management scheme for allowing easy access to the data and for rendering the results of the online queries in a user-friendly manner, also by means of special graphical interfaces.

In the present paper both aspects are presented and discussed. Special attention is given to the prosodic analysis of the database, in terms of the methodologies followed in the labelling phase as well as the consequent strategies adopted in the implementation of software tools for prosodic analysis.

2. Checking the database coherence

A very important step before putting the data into a DBMS is to check the consistency of temporal alignment of labels across all levels of annotation. This is crucial in the case of the AVIP corpus, since time is the unique basic reference for all data, and therefore an high level of confidence on time markers alignment is a stringent

condition. Even though SegWin, the software tool developed during the AVIP project for labelling the corpus, was specifically designed also for checking temporal alignment across the levels during the annotation phases (Refice et al. 2000), an offline software tool named DB-Checker has been implemented which checks the temporal alignments across the different tiers and reports the detected discrepancies. Some specific and systematic errors are automatically corrected by the system; when an automatic correction cannot be performed with the needed confidence, then human transcribers intervene to manually correct them, and this procedure can be iterated as many times as needed, until DB-Checker does not detect any other alignment error.

3. Prosodic labelling in the AVIP corpus

One of the main challenges in building and labelling a corpus like AVIP is represented by the inadequacy of phonological and phonetic descriptions of Italian regional accents, especially at the prosodic level. The strategy used to cope with such a problem has basically consisted in adopting an iterative process of labelling, such that a description for each variety could be achieved by successive refinement stages without losing intermediate stages information (Refice et al., 2000). For the nature of the descriptions in itself and for the amount of phonological knowledge already available, labelling at the segmental levels using such procedure has proved to be less difficult than prosodic labelling, for which two different approaches have been discussed and followed within the project: a) a more “phonologically oriented” autosegmental-metrical ToBI-like one (Ladd, 1996), and b) a more “phonetically-acoustically oriented” one, inspired to the INTSINT model (Hirst & Di Cristo, 1998).

The present paper reports and describes intonation labelling procedures and software tools developed for the Bari variety of Italian, whose intonation phenomena in the AVIP corpus have been labelled according to an autosegmental-metrical ToBI-like approach, since for it a preliminary tonal inventory has been proposed in previous work, based on analysis of spontaneous Map Task dialogues of the same nature as the ones constituting the AVIP corpus (cfr. for example Grice & Savino, 1995; 1997; Savino, 1997; 2001). In doing this, we yet adopted

specific strategies, along the lines stated above (“iterative process of labelling”), for a more detailed description of tonal events whose phonological identity has been only preliminarily defined for this variety of Italian. In order to verify such phonological hypotheses over a larger database, during the first labelling stage a more “phonetically-oriented” description with respect to the typical ToBI scheme (Beckman & Ayers Elam, 1994) has been applied: this was considered as necessary especially in order to reliably determine the phonetic realisations of tonal categories. For example, in case of assumed bitonal pitch accents two separated labels have been used, one for temporally positioning the leading/trailing tone, and the other one for marking the position of the starred tone on the F0 curve. Therefore, typical Bari Italian interrogative L+H* pitch accents have been labelled by positioning the label ‘L+’ on the local F0 minimum preceding the peak, and the label ‘+H*’ on the F0 peak. In this way, the symbol ‘+’ preserves the information that the two labels are constituents of a bitonal pitch accent., whereas its position (left/right of the label) bears the information about the function of the tone (leading/trailing): so the bitonal pitch accent H*+L is “composed” of the two labels ‘H*+’ and ‘+L’.

The same strategy is adopted for labelling intonation boundary phenomena: unlike the typical ToBI scheme, where a complex final rising F0 movement is marked by positioning, for example, a L-H% label at the highest F0 point on the curve corresponding to the rightmost edge of the phonological utterance (being the phonetic realisation of that event already determined in English, Beckman & Pierrehumbert, 1986; Beckman & Ayers Elam, 1994), we put a temporal marker at the point of the curve where the local minimum occurs, and a separated H% label for marking the highest F0 point at the end of the segment curve (and corresponding to the IP boundary). A representation of the labelling scheme described above is shown in Figure 1.

Truncation phenomena, which have been attested for this variety of Italian (Savino, 1997; Grice et al., forthcoming), are explicitly marked by putting the not fully realised boundary tone symbol in parentheses: for example, final falling truncation is represented by positioning a complex label ‘L-(L%)’ at the end of the falling bit of the F0 curve.

It is also to be noted that in the inventory of Bari Italian pitch accents a special ‘n’ flag is used for labels identifying nuclear accents, as it has been ascertained that in this, as well as in other varieties of Italian, nuclear accents cannot be defined positionally as in English (Grice et al., forthcoming), and therefore a special label is necessary. Moreover, all uncertainties are labelled, by using the same conventions established in standard ToBI (Beckman & Ayers Elam, 1994).

Unlike the tiers at the segmental levels, which are all temporally aligned according to the hierarchical scheme described in (Refice et al, 2000), in the AUT(osegmental-metrical) ToBI-like annotation level adopted in AVIP the only alignment required is that between the IP tonal category and the segmental levels, i.e. a boundary tone label L% or H% is always temporally aligned with the rightmost edge of a word (and, consequently, of a phonetic segment in the relating phonetic tier). Accordingly, the above mentioned program DB-Checker verifies also the consistency in temporal alignment of the IP labels in the

AUT(osegmental-metrical) tier with the segmental tiers.

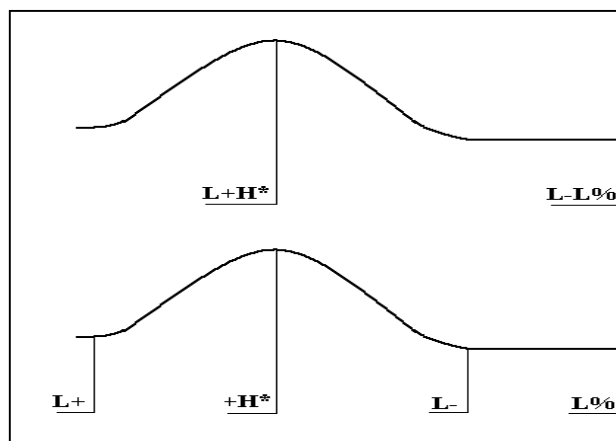


Figure 1. Exemplifying scheme of Bari Italian intonation labelling procedure in the AVIP corpus using a more “phonetic” approach (bottom) with respect to standard ToBI (top)

4. The AVIP-Analyser

It is not rare that available speech corpora are not supported by specific programs/routines for making online computations on the information extracted from the database. While this is not a problem for users who have also programming skills, and therefore can develop themselves the needed programs/routines for the analysis, such situation tends to exclude, or at least to strongly limit, other users from the possibility of querying and analysing a speech database without suitable interfaces, especially graphical ones. Starting from these considerations, we inserted the AVIP data into a Data-Base Management System (namely, Microsoft SQL-server), and implemented a specific user interface which avoids the need of writing complex logical expressions for querying the database (query-generator). By using our system, called AVIP-Analyser, it is possible to perform statistical analysis both at segmental and suprasegmental (more specifically, intonation) levels. Figure 2 shows a snapshot of the screen appearing when the user’s choice is the intonation analysis. As it can be noted, the possible selections (irrespective of the level of analysis) include:

- the map used during the Map Task session (A, B...);
- the variety of Italian (Bari, Napoli, Pisa);
- the recording session number;
- the turn number;
- the role of the speaker (Instruction Giver or Follower);
- sex and (range of) age of the speakers.

Moreover, by “enlarging” the usual way of Data-Base Management Systems to operate with strings and/or numbers, our system allows also the graphical rendering of the results of specific queries, performing, when necessary, online computations by means of specialised routines automatically linked and executed by the query generator. These functionalities are particularly useful in analysing prosodic aspects of the speech data, especially as far as F0 parameter is concerned. At a very basic level, these functionalities allow to check correctness and consistency of the database, for those aspects which

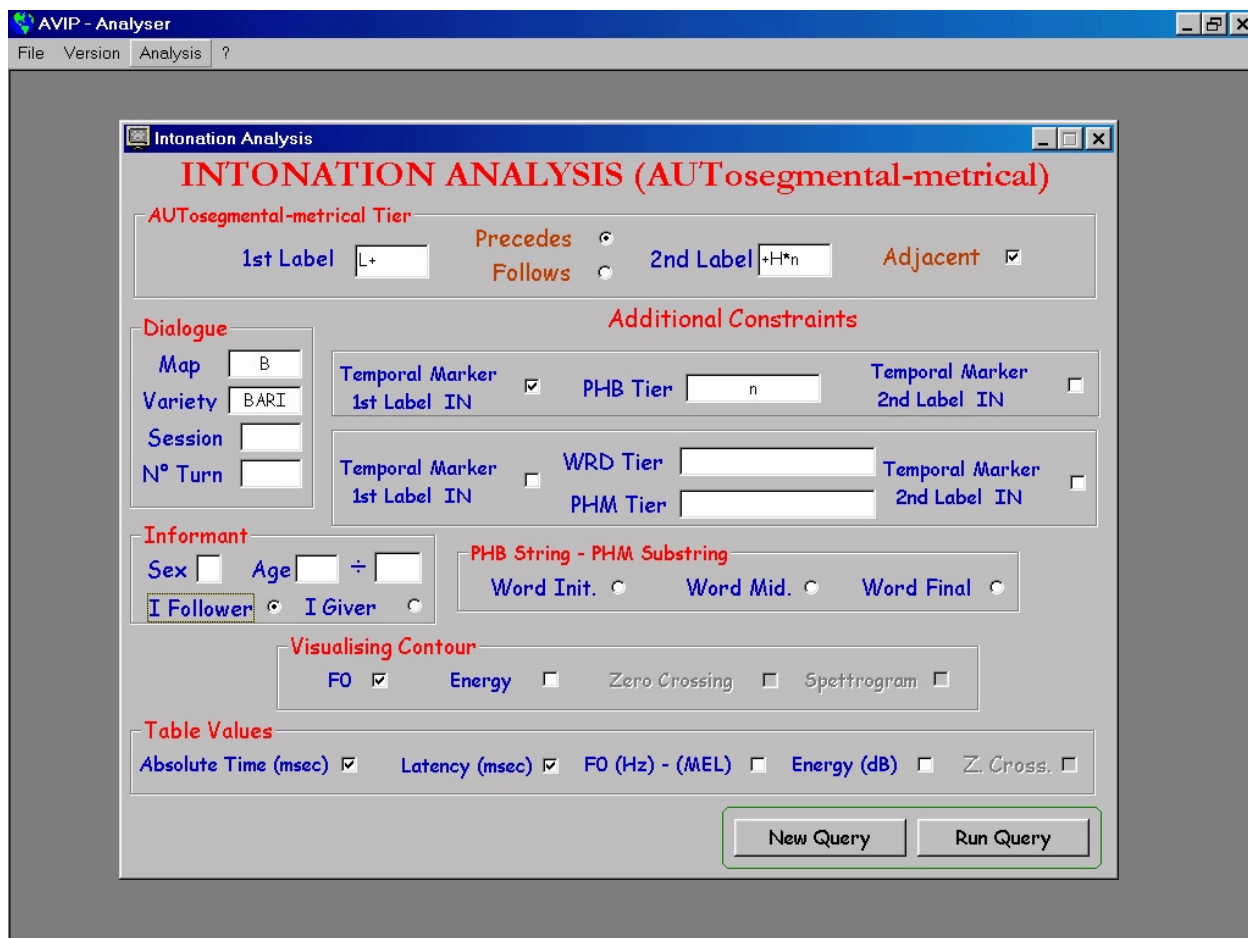


Figure 2. Snapshot of the intonation query panel of AVIP-Analyser

cannot be checked by using the DB-Checker, namely:

- the correct positioning of an intonation label with respect to the conventions adopted;
- the consistency of intonation labelling in terms of phonological category assigned
- errors in pitch tracker estimation values which deeply influence statistical results on data.

4.1. Tonal alignment/association

Basing on the “phonetically-oriented” strategy adopted for intonation labelling in the corpus, the AVIP-Analyser has been designed and implemented to perform also specific queries for statistically verifying the phonological hypotheses concerning Bari Italian tonal inventory within an autosegmental-metrical ToBI-like framework. In fact, besides the standard queries for obtaining, as an example, absolute F0 or time values for a given category (label), the system allows the computation of the latency occurring between two adjacent tonal events, for example:

- latency between the leading/trailing tone and its relating starred tone
- latency between a pitch accent (starred tone) and an adjacent intermediate phrase boundary/accent.

Since latency may be influenced by the phonetic characteristics of the associated segments, the AVIP-Analyser allows to perform statistics posing also constraints on the position of a given intonation label in relation to the segmental tiers. For example, one can get

the latency between L+ and +H* where L+ is realised in a nasal consonant (PHB refers to the PHonetic Broad transcription level in the database), as it is shown in Figure 2. Additional constraints may be added concerning also the relative position of the segment within a word (in Figure 2, WRD and PHM refer to the word-by-word segmentation in the AVIP database, orthographic and phonemically transcribed levels respectively).

By means of this kind of statistical analyses is therefore possible to determine the phonetic realisations of any tonal category for any given type of utterance.

4.2. Consistency of intonation labelling

In labelling a large database, the risk of inconsistencies throughout the corpus may be quite high. The AVIP-Analyser functionalities are intended as an aid in coping with the problem of pinpointing such cases, since they allow to obtain also, as output of a query, a graphical rendition of all the F0 curves related to the queried events, each of them being suitably highlighted within the F0 track of the whole reference utterance, as it can be seen in Figure 3, Figure 4, and Figure 5. In particular, Figure 3 shows results of F0 values (in Hz and Mel scales) of the queried L+ and +H*n adjacent labels, where for each token a visualisation of the F0 curve is also possible. Figure 4 shows latency values and visualisation of relating F0 [L+ +H*n] segment on the curves for one selected token, and Figure 5 shows a graphical rendering

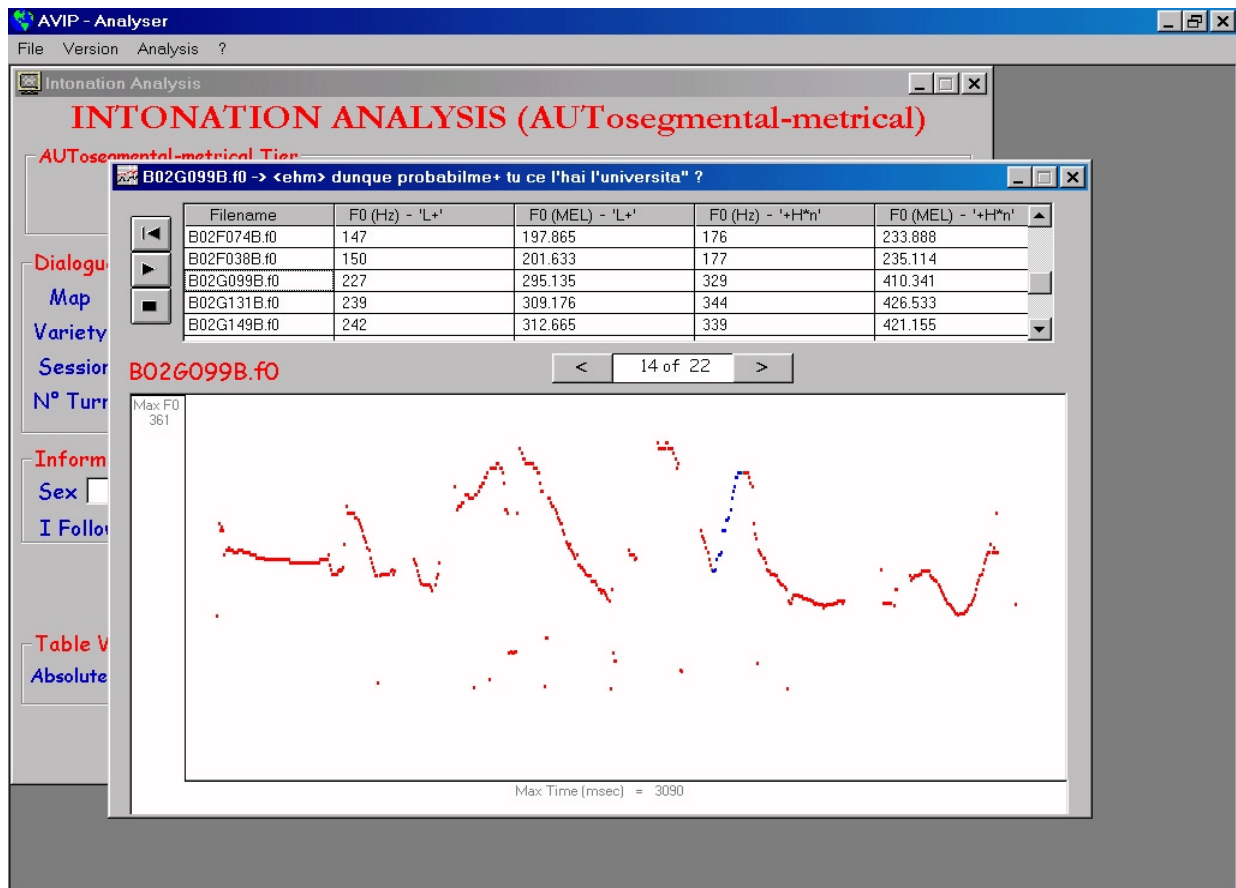


Figure 3. F0 values (in Hz and Mel scales) of queried adjacent targets L+ and +H*n with visualisation of relating F0 segment on the curve for one selected token

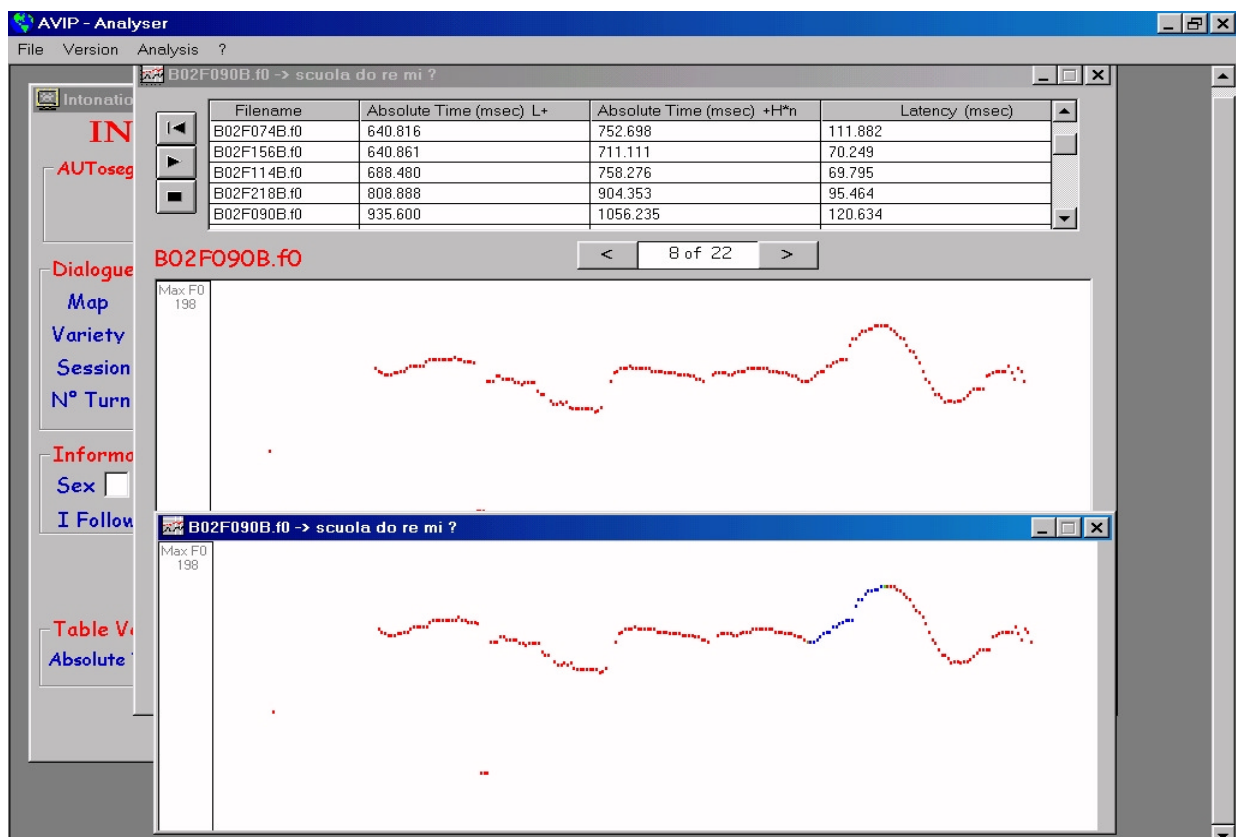


Figure 4. Latency and visualisation of relating highlighted [L+ +H*n] F0 segment for one selected token

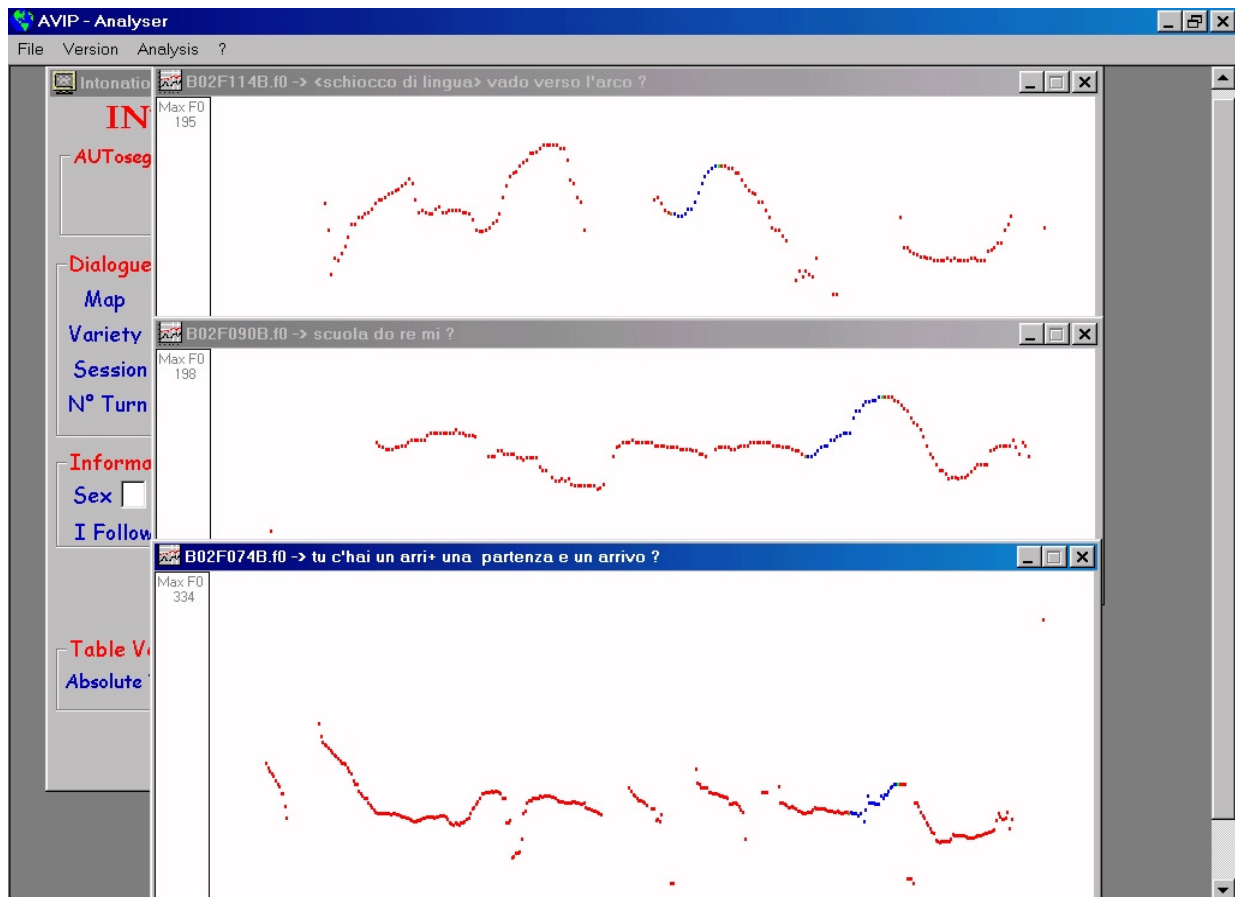


Figure 5. Visual rendering of queried adjacent [L+ +H*n] F0 segments for a number of selected tokens

of a possible comparison among different tokens in the same query. Moreover, for each result of a given query, the play function allows to listen to the target utterances. Such functionalities are therefore useful not only for checking the consistency of the assigned labels, but also for the statistical assessment of the hypotheses made upon the phonological categories of Bari Italian intonation.

4.3. Intonation modelling

One of the problems not yet appropriately solved in intonation analysis and labelling of spontaneous speech, consists in dealing with cases of more or less strong F0 perturbations, coming from a number of well-known acoustic/phonetic phenomena (presence of unvoiced segments, creaky voice, etc.). Several solutions for modelling F0 curves have been proposed (for example Hirst et al., 2000), but the problem of establishing the correct anchor points for determining the appropriate contours still remains.

Given that the temporal positioning of the intonation labels in our corpus may be sometimes questionable, due to the above mentioned F0 perturbations, we need an iterative approach in order to be able to better determine such anchor points, using the preliminary markers as a first hypothesis. In this respect, the system AVIP-Analyser may be used in building up hypotheses on the temporal markers, and save them as “dynamic models” intonation tiers to which make reference for performing statistical

analysis upon the still real data of the corpus.

As a matter of fact, our tool allows us to perform several different stylisations upon a given F0 curve, put any “modelling intonation anchor point” on that curve and store all the corresponding data as separated “model” files, which can be appropriately selected, in place of the originals, in any generated query. In other words, we can create a “model” intonation tier, and perform any on-line statistics based on that model, but relating to all the other real data, check the consistency of the model with respect to a given phonological hypothesis, and so on iteratively.

5. References

- Albano Leoni, F., Paoloni, A., Refice, M., Rinaldo, M., Savino, M., Sobrero, A., 1998. CLIP: Corpus della Lingua Italiana Parlata (Corpus of Spoken Italian). In A. Rubio, N. Gallardo, R. Castro, A. Tejade (eds.), *Proc. LREC 1998* (1 International Conference on Language Resources and Evaluation), Granada: La Gioconda, 1:503-506.
- Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., Weinert, R., 1991. The HCRC Map Task Corpus. *Language and Speech* 34: 351-366.
- Beckman, M. E. and G. Ayers Elam (1994). Guidelines for ToBI labelling. version 3.0, March 1997. Ohio State University.

- Beckman M. and J. Pierrehumbert (1986). Intonational structure in English and Japanese. *Phonology Yearbook*. 3. 255-310.
- Grice, M. & Savino, M., 1995. Low tone versus 'sag' in Bari Italian intonation: a perceptual experiment. In K. Elenius & P. Branderud (eds.), *Proc. ICPHS 95* (Proceedings of the XIII International Congress of Phonetic Sciences), Stockholm: Stroemberg's Grafiska, 4: 658-661.
- Grice, M. & Savino, M., 1997. Can pitch accent type convey information status in yes-no questions?. In K. Alter, H. Pirker, W. Finkler (eds.), *Proc. of the ACL workshop "Concept to Speech Generation Systems"*, Madrid:UNED. 29-38.
- Grice, M., D'Imperio, M., Savino, M., Avesani, C., forthcoming. Strategies for intonation labelling across varieties of Italian. In Sun-Ah Jun (ed.), *Prosodic Typology: an Approach through Tone and Break Indices*. Oxford: Oxford University Press.
- Hirst, D. & Di Cristo, A., 1998. A survey of intonation systems. In D. Hirst & A. Di Cristo (eds.), *Intonation Systems*. Cambridge: Cambridge University Press.
- Hirst, D., Di Cristo, A., Espesser, R., 2000. Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (ed.), *Prosody: theory and experiment*. Dordrecht: Kluwer, 51-87.
- Ladd, R.D., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Refice, M., Savino, M., Altieri, M., Altieri, R., 2000. SegWin: a tool for segmenting, annotating, and controlling the creation of a database of spoken Italian varieties. In M. Gavrilidou, G. Carayannis, S. Markantonatou, S. Piperidis, G. Stainhaouer (eds.), *Proc. LREC 2000* (II International Conference on Language Resources and Evaluation), Athens: National Technical Univ. of Athens Press, 3:1531-1536.
- Savino, M. 1997. *Il ruolo dell'intonazione nell'interazione comunicativa. Analisi strumentale delle domande polari in un corpus di dialoghi spontanei (varietà di Bari)*. unpublished Ph.D. thesis, Politecnico/Università di Bari, Italy.
- Savino, M., 2001. Non-finality and pre-finality in Bari Italian intonation: a preliminary account. In P. Daalgaard, B. Lindberg, H. Benner, Z.-H. Tan (eds.), *Proc. Eurospeech 2001*, Aalborg: Kommunik Grafiske Losninger, 2: 939-942.