Data-Driven Intonation Teaching
An Overview and New Perspectives

Valentina De Iacovo
Università degli Studi di Torino, Italia

Antonio Romano
Università degli Studi di Torino, Italia

Abstract  Studies on the teaching of intonation are not new. They have actually represented an area of interest since the beginning of the 20th century (Jones 1909; Palmer 1922) up to the latest works by authors following different approaches such as Wells (2006), Canepari (2008), Busà (2012). More recently, there has been growing interest in computer assisted learning (Chun 1998, 2013) and applications in commercial products (Cazade 1999). In this contribution we provide a preliminary study which considers the degree of correlation between the intonation curves of L2 English learners and English native speakers in order to explore variation and to evaluate the potential proximity between different samples. The ultimate goal is to test a set of techniques designed to optimise a computer assisted program for second language intonation learning which have been developed for prosodic variation assessment (Cauvin 2017, De Iacovo 2019).

Keywords  Computer-assisted language learning, Italian and English intonation, L2 teaching.

Summary  1 Intonation Teaching: An Overview. – 2 The Assisted Teaching of Foreign Languages’ Intonation. – 3 A Pilot Study. – 4 A Method Based on the Prosodic Distance Assessment. – 5 Conclusions.
1 Intonation Teaching: An Overview

When we refer to the study of speech, the prosodic aspects of a language occupy a prominent position. Among the most relevant features described, rhythm and intonation play a central role. The earliest studies on intonation teaching date back to the beginning of the 20th century. A considerable number of scholars (see, among others, Jones 1909, fig. 1; Palmer 1922) focused on the prosodic aspects of foreign language teaching, considering these elements pivotal “to give the final touches to a good pronunciation” (see Jones 1922, IV). Through their personal experiences, scholars realised the importance of teaching not only pronunciation but also intonation. By means of curves, dots, notes, they endeavoured to sensitise students to all those suprasegmental phenomena (such as pitch, duration, etc.) which contributed to the reproduction of what Pierre Delattre would later call “the salt of an utterance” (cf. Delattre 1966b, 81).

![Figure 1: Examples of English (which is your nearest post office?) and French (qui te rend si hardi de troubler mon breuvage?) utterances described as continuous curves on a musical score (Jones 1909, 17, 41)](image1.png)

![Figure 2: Examples of American English utterances associated with two intonation patterns. Syllables are raised or lowered on a line in order to reproduce their pitch within sentences with a different meaning (Bolinger 1978, 479)](image2.png)

![Figure 3: Example of an utterance (I thought it was one and five!) in American English with phonetic transcription and raw indications about presumed pitch movements associated to prominent syllables (Palmer 1922, 53, ex. 141)](image3.png)
The British school employed different methods of representation on the description of English, French and German intonation (as illustrated in fig. 1) within a superpositional view of this phenomenon. Nevertheless, some authors (Palmer 1922; O’Connor, Arnold 1961; Gussenhoven 1984; Cruttenden 1986) suggested a simplification based on a limited number of levels and configurations, whereas, various other authors would gradually shift towards linear models.¹

The French school has also contributed to the identification of basic melodic contours for different European languages, namely Delattre (1966b),² and, more recently, Hirst (1983), Hirst, Di Cristo (1998) and Wells (2006). Hirst defined a model where intonation is the result of different prosodic layers: the accentual group, the tonal unit (TU) and the intonation unit (IU). Other models were suggested by Martin (1987), Rossi (1999) and, more recently, Lacheret-Dujour (2001) who, inter alia, shed light on the role played by information structure.

¹ Various works on English intonation (Halliday 1967; Crystal 1969; Couper-Kuhlen 1986; Hirst 1998; Wells 2006) assumed functional but still phonetic models, whereas some others (’t Hart, Collier, Cohen 1990) adopted a more perceptive view. A general phonological framework is offered nowadays by the Autosegmental-Metric model (Ladd 1996). Details about the differences between these models in respect of intonation teaching are in Romano, Giordano 2017.

² He described 10 fundamental intonations as global movements on a four-level scale.
Italian prosody has a strong diatopic connotation. In terms of didactic approach, this raises the following questions: which kind of intonation should we teach? How can we show the prosodic diversity? First studies on Italian geoprosodic variation date back to Panconcelli-Calzia (1939) whereas several authors, such as Canepari (1985), Chapallaz (1960) and later Rossi (1998) and Grice et al. (1999, 2005), worked on less geographically connotated data. More recently, works have extended their focus to include the regional description (and differentiation) of major prosodic patterns. Also, aspects such as the relation between syntax and intonation and $f_0$ contours and pragmalinguistic functions have been investigated (De Iacovo 2019).

Despite growing prominence in L2 didactic textbooks (thus confirming the main role played by prosodic events), these phenomena have not received a fair theoretical framework as yet, and they are still presented in very general terms (an early attempt can be seen in the good practice encouraged by Guimbretière 1994).

In some cases, teachers have adopted practices more focused on the pragma-linguistic organisation based on more theoretical advances on speech structuring.³

2 The Assisted Teaching of Foreign Languages’ Intonation

As suggested by many scholars (Boureux, Batinti 2004; Trouvain, Gut 2007; De Marco, Sorianello, Mascherpa 2014), the role played by prosody in the pronunciation accuracy of a foreign language is crucial. Despite the considerable literature focusing on prosody is evident, its classroom application is still insufficient, with suprasegmental features representing one of the less evaluated parts in L2 teaching (Busà 2012). It is important to offer students specific prosodic models of a foreign language by means of selected sentences showing pauses, stresses and typical intonation patterns. Previous models were reported by Chun (1998), who introduced the Visi-Pitch software, where students can compare the intonation curves of their uttered sentences with the same ones produced by a mother tongue speaker.⁴ Similar studies by Cazade (1999) and Delmonte (1999) are specialised in prosodic tutorials by means of self-learning exercises which visually highlight the differences between the prosodic patterns of English and Italian [fig. 6].

³ For Italian, it is worth mentioning Cresti 1999 and, more generally, previous works based on studies by J.L. Austin, J.R. Searle, C.B. Benveniste.
⁴ First applications of this method can be found in James 1976 and Fischer 1986.
Figure 6  Examples of visual matching between pitch curves extracted from students’ utterances and the corresponding curves in similar utterances by native speakers. The sentences are: a) Qu’est-ce qu’il fait? (Fischer 1986); b) Can you manage? (Delmonte 1999)
Thanks to technological advances, research could make rapid progress and other prosodic correlates have been investigated (for languages such as English and French, see Frost 2010; Frost, Picavet 2014). Simultaneously, other authors (Peperkamp, Dupoux 2002) started developing new products focusing, for example, on stress and intended for “stress-deaf” languages (such as French). Among the other phono-didactic oriented products, it is worth mentioning the software SWANS and the prototype Sounds right [fig. 7]. Other experiments – in progress – concern the Italian intonation learned by Chinese students and can be found in ww, Pettorino (2011), Pettorino, Vitale (2012). Other authors (Busà 2012) reported the lack of a compactness in terms of methods and tools available for the teaching and learning of prosody. Also, it is now commonly accepted that being aware of how prosody works in one’s native language may sensitize the student to establish the differences between the two languages (Frost, Picavet 2014; Romano, Giordano 2017).

Figure 7  Example of teaching visual display offered from the software Sounds Right (Péchou, Stenton 2001)
3 A Pilot Study

On the basis of these premises, we decided to carry out a series of experiments in order to compare the prosody of English L2 students and 4 teachers whose mother tongue was English. The first recordings of 16 students took place in a school, whereas in a second step recordings were collected in a sound-proof booth of the Laboratory of Experimental Phonetics “Arturo Genre” of the University of Turin and consisted in 20 sentences read by 10 native Italian students (mean age = 23 years old). Each speaker was given the list of 20 sentences and was asked to read them once at normal speed, as if they were saying the sentences in spontaneous speech. If the sentence was wrongly pronounced, they have been asked to repeat it. An instrumental analysis was later conducted using the software Praat in the Laboratory of Phonetics “Arturo Genre” of the University of Turin. Utterances were firstly manually segmented and labelled by isolating the vocalic portions [fig. 8], then acoustic values of $f_0$, duration and energy were extracted and later compared in terms of prosodic distance by means of a correlation matrix obtained by a procedure developed within the AMPER project. Figure 9 shows a clustering tree of all speakers based on their relative prosodic proximity: for example, students 4 and 9 (named s4 and s9) are the closest to the teachers’ prosody (respectively named mra, pau, nan).

---

5 A preliminary study, whose premises are described in Romano, Giordano 2017, is based on the comparison of 50 sentences read by 15 English learners who were native speakers of Italian aged between 16 and 17 years old and an English mother tongue teacher. For this study, sentences were firstly read by the teacher and later repeated by the students.

6 For example, by adding pauses, mumbling or changing words.

7 For a general reference, see De Iacovo 2019.
Figure 8 Example of vowel segmentation of the sentence she’d met him before then, hadn’t she?

Figure 9 Cluster of the sentence she’d met him before then, hadn’t she? for all speakers based on $f_0$, duration and intensity.
The dendrogram in figure 9 shows a first distinction of two main groups, where the first one (on the left) group together the English teachers and various students (u4, u5, u6, s4, s5, s9, s11).

Another possible graphic representation is the heatmap [fig. 10], which shows the cluster proximity in terms of colour/greyscale: red represents the maximum correlation, blue the minimum. The teachers’ performances (mra, nan, pau) are well correlated to student 4 (s4), while the fourth teacher (cnd) presents a correlation with student u6 and u4. Different students (u2, u3, u8, u9, u12, s2, s3, s6, s7, s12, s13, s14, s15) seem to have a strong correlation (upper right red rectangle) with each other, which may be interpreted as the realisation of a very similar prosodic pattern for all of them.

Figure 11 shows the correlation among the speakers by means of a phylogenetic tree where each group (squared boxes) is associated to the correspondent intonation curves: for instance, students 4 and 9 (s4 and s9) or students 5 and 11 (u5 and s11) present very similar curves, whereas sentences by mra, nan and pau confirm their strong similarity.

The prosodic scheme of these students was already well correlated to the teachers’ one in the previous sentence [fig. 9].
4 A Method Based on the Prosodic Distance Assessment

When referring to L2 teaching, we have several solutions to apply the methods mentioned above with regard to the assessment of learners via the evaluation of other competencies (for example, we can judge the intonation by asking the student to answer a question and repeat it). We attempted to investigate a method based on the comparison of prosodic curves by means of a correlation matrix. This method can be broken down into several phases and is intended for a potential CALL (Computer-Assisted Language Learning, see Levy 1997) application. During the first phase, the user sees the text of a question and listens to it from a pre-recorded file. Secondly, a list of n possible answers appears, and the user has to choose the right one, by clicking on a button. If the answer is wrong, the system suggests making another attempt, until the answer is correct. At this point, the user has to record his answer and a button appears to record his voice. He receives some practical indications about how to record his voice when reading the answer (particular attention should be given to potential pauses or words’ addition/deletion, for example). The user presses stop in order to save the soundfile and the system compares the recording with the learner’s answer and confirms if the answer is correct. If the answer differs for one or more words, the system asks the user to repeat the answer. The last recorded answer is...
then automatically segmented into vocalic portions and $f_0$, duration and intensity values are extracted from the segmented file using a script. The extracted data is then compared with the production of native speakers via a minimum distance algorithm. At this point, we are seeking to develop a method to show the intonation curves of the native speaker and the user, giving a score and highlighting potential parts which do not match. If the score value is lower than an established minimum threshold, the system suggests a new attempt.

In terms of didactic enhancements, this kind of interaction serves a number of purposes: the recorded corpus is available for research purposes on the variation of L2 intonation; plus, the comparison with the pre-recorded utterances allows the user to improve the pronunciation and intonation of at least a sentence and to highlight the mismatches with the pronunciation of native speaker; furthermore, the fact of displaying the curves allows the user to pay attention to the prosodic organisation of the target language; finally, the aligned written text and the oral utterances together sensitise the user to spot the differences between oral and written language.

5 Conclusions

This article has aimed to give an overview on the data-driven intonation teaching. After the introduction, which shows the main pioneering studies on the teaching of prosody, we focused on the most recent applications involving the CALL research. Prosody represents an important challenge when learning a second language and nowadays we have all the instruments required to improve the students’ linguistic skills. The present study aimed at giving a possible path to measure the prosodic distance and to sensitise the student by giving him graphical feedback. There remain issues: a generalised theoretical model covering different languages has not been addressed as yet; a quantitative approach is also necessary, aiming at covering all potential linguistic variations, in terms of communication schemes and information patterning. Furthermore, the process of data analysis should also include data from various sources, including physical parameters. Finally, other linguistic dimensions have to be examined: discourse level, focus, diatopic variation, extra-linguistic features. With the right knowledge and instruments, this should lead to more refined applications for the L2 prosodic training.

9 Scripts are taken from the AMPER project (see Romano, Contini, Lai 2014).
Bibliography


