DOI: 10.17469/O2108AISV000015

VALENTINA DE IACOVO, MARCO PALENA, ANTONIO ROMANO

Evaluating prosodic cues in Italian: the use of a Telegram chatbot as a CALL tool for Italian L2 learners

Comparing the pronunciation of learners of a foreign language (FL) with utterances of native speakers is receiving attention thanks to the increasing number of applications arising in the field of computer-assisted teaching (Cazade, 1999) and computer-assisted language learning (Levy, 1997). At the same time, language-learning studies on prosodic inter-speaker variation bring to light a rhythmic-intonational variability that should not be reduced to few acceptable patterns but be part of the learner's linguistic background. It still seems open to debate how best to make explicit to the learner his degree of prosodic competence on the basis of acoustic features (f_0, intensity, duration) and other correlates (syllables, lexical accent, speech rate, pauses). In this study we present a chatbot as a proactive learning support to improve the oral skills in Italian L2.

Keywords: L2 intonation, Italian prosody, oral production, CALL, educational chatbot.

1. *Introduction*

The prosodic features of speech reveal important information such as the enunciative mode, the speaking style, the attitude and the regional or social connotations of the speaker, and are therefore essential for successful interpersonal communication (Hirst, 1983; Cruttenden, 1986). However, in many contexts related to foreign language teaching (FL), metaprosodic reflection is not always satisfactory for the learner because it is not dealt with comprehensively. Referring to the role played by prosodic features such as intonation, rhythm, duration, and focus to explain communicative differences in meaning may instead prove indispensable in triggering an awareness in the learner in recognising and reproducing FL speech (Cresti, 1999; Chun, 2002; Frost, Picavet, 2014). Moreover, for several decades now, technological developments in speech analysis have opened the way to new perspectives on communicative interaction that can be used in language teaching (Chun, 1998; Cazade, 1999; Busà, 2012; Romano, Giordano, 2017), and have also led to the emergence, from the very first studies conducted, of criticism towards possible solutions for the explanation of the degree of competence achieved, for example through the use of graphic representations (James, 1976; De Bot, 1983; Martin, 2010). The teaching of oral FL allows just that kind of metaprosodic reflection (Trouvain, Gut, 2007) thanks to which students can experiment the variety of enunciative combinations and associate their meaning on the basis of specific functions (modal, syntactic, informative, emotional) in order to refine their oral communication skills (Delattre, 1966; Cresti, 1999; De Iacovo, Romano, 2019).

In Italian it is also necessary to take into account the different intonational choices on the basis of the regional or local variety used (Canepari, 1983, 1985; Bertinetto, Magno Caldognetto, 1993; Sorianello, 2006) included for some time now in the research on L2 intonation (De Meo, Pettorino, 2012; De Marco, Sorianello & Mascherpa, 2014) that allows the learner a greater choice among prosodic models. Finally, we find it interesting to focus on the different possibilities of realization typical of oral language and how they should be taken into account in the context of automatic evaluation and recurrence of intonation patterns.

On the basis of these considerations, our work is structured as follows: after setting up the chatbot's steps for prosodic evaluation, we recorded a set of ten complex sentences constructed to give an adequate representation of the most frequent intonational solutions of a formal speech on technical-scientific contents, but also including some everyday expressions by native Italian speakers to be used as a comparison with Italian learners. The results obtained thanks to the comparison between the sentences produced by Italian speakers and learners thus provide a first cue to reflecting on some considerations that have emerged: in what way do the values of the acoustic parameters extracted by the chatbot reflect a perceptual evaluation? (Munro, Derwing, 1999) What are the most recurrent prosodic patterns used by Italian speakers and students? What is the degree of reliability of the automatic segmentation system with respect to the results returned? (Lacheret-Dujour, 2001).

2. *Chatbot structure and functioning*

In this section we illustrate how the chatbot has been structured as an assisted teaching tool for the assessment of prosodic features. The adoption of such an instrument has several advantages. First of all, it provides a text-based mode of interaction to which the public is already largely accustomed thanks to the popularity of instant messaging applications such as *Whatsapp* or *Telegram*. This makes these tools easier to use than dedicated applications or Learning Management Systems (LMS) such as *Moodle* or *Blackboard*, whose main obstacle to usability is the need for users to adapt to an unfamiliar interface. Moreover, a chatbot allows for interaction with learners in a structured and asynchronous way: structured because it is organised according to predefined schemes, aimed at reducing ambiguities; asynchronous because it is not bound to continuous time intervals but can be used freely according to one's own time availability. The potential of chatbots as teaching aids has been investigated for a few years now (Pereira, 2016; Fernoagă, Stelea, Gavrilă & Sandu, 2018) and several recent studies have focussed on their pedagogical (Cheng, Lau, Lam, Zhan & Chan, 2020) and prosodic assessment (Lezhenin, Lamtev, Dyachkov, Boitsova, Vylegzhanina & Bogach, 2017).

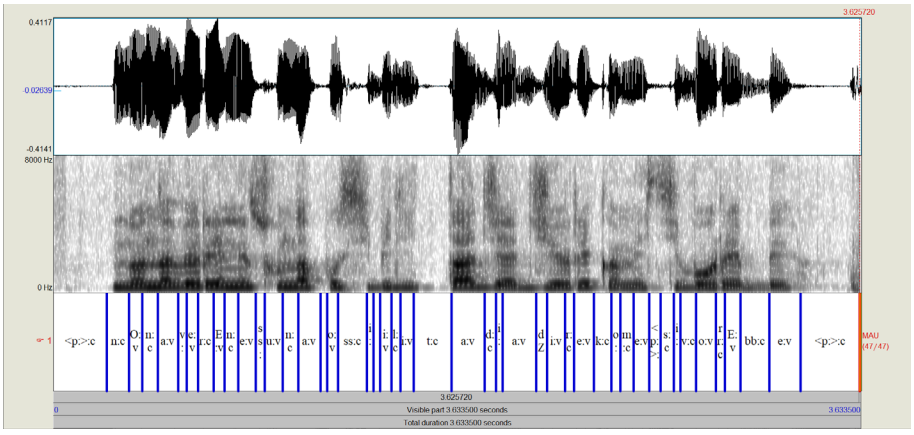
Implemented within the instant messaging application *Telegram*, the chatbot (or bot) developed involves interaction with the user through questions and answers based on the assessment of technical knowledge. In particular, it directs to the learner a series of closed-ended questions (quizzes) that can be of a general nature

of linguistic comprehension or linked to a particular subject area (e.g. technical or scientific). When the correct answer is found, the learner sends a statement of the answer by voice message. The received utterance is then automatically processed by the bot in order to obtain an evaluation of the intonation level of the speaker. This processing can be summarised in the following steps:

1. Pre-processing
2. Phonetic segmentation
3. Extraction of fundamental frequency values
4. Intonation analysis

The first step in the processing chain has a dual purpose: to guarantee a minimum quality level for the input audio signal and to prepare it for further processing. The audio is first converted to single-channel wav format, resampled (if necessary) to 48 kHz and cleaned of background noise. The average amplitude value and signal-to-noise ratio of the resulting audio are then estimated. If the estimated values are below certain predefined thresholds, the chatbot prompts the learner to record a new utterance in a less noisy environment and/or by speaking in a higher tone of voice. The speech signal is then subjected to segmentation using the *WebMAUS* Basic web service (Kisler, Reichel & Schiel, 2018). This service, taking as input the speech signal and the orthographic transcription of the utterance, returns a segmentation into words and phonemes based on the method described in Schiel (1999). The phonetic segmentation of the utterance, provided in TextGrid format to facilitate subsequent processing using *Praat* software (Boersma, Weenink, 2018), is then processed by labelling individual phonemes as vowels or consonants (Fig. 1). A *Praat* script is then invoked to extract the fundamental frequency (f_0) values of the previously identified vowel phonemes, thus obtaining the intonation curve of the utterance (for the segmentation and value extraction model, see Romano, Contini & Lai, 2014).

Figure 1 - Praat *screen (wav and Textgrid) showing the phonetic segmentation made by* WebMaus *for the utterance "Non avere nessuna possibilità di agire come si vorrebbe"*



The analysis of intonation is carried out by comparing the intonational curve of the learner's utterance with the f_0 traces of the corresponding utterances of native speakers, previously collected and evaluated using the same automatic procedure. The comparison is made by calculating a correlation measure (Moutinho de Castro, Coimbra, Rilliard & Romano, 2011) that compares for each sentence three points (initial, central, final) of f_0 of each vowel segment identified by *Maus*. Such a measure is known as the Pearson correlation coefficient:

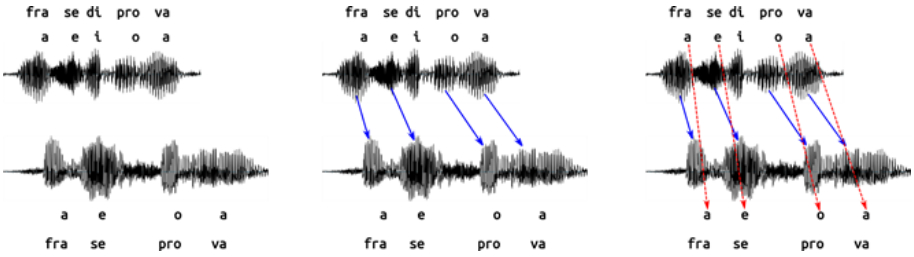
(1)
$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \text{ with: } -1 \leq \rho_{XY} \leq 1$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y)$$

where X and Y are two data series of n values each, corresponding to the f_0 traces, the values μX and μY represent the means of these data series and the values X and σY represent their standard deviations. The Pearson coefficient is a measure of linear correlation and its value ranges between -1 and 1 (or in percentage between -100% and 100%). A positive value indicates linear correlation between the two data series, meaning that most pairs of f_0 values tend to be simultaneously greater than, or simultaneously less than, their respective means. A value of 1 indicates perfect linear correlation, meaning that each pair f_0 values increases or decreases accordingly and always with the same proportions. Conversely, a negative value indicates anti-correlation, meaning that the pairs of f_0 values tend to lie on opposite sides of their respective means. A value of -1 indicates perfect anti-correlation, meaning that for each pair of f_0 values, as one increases the other decreases and always with the same proportions. A value of 0 implies that there is no linear correlation between the two data series.

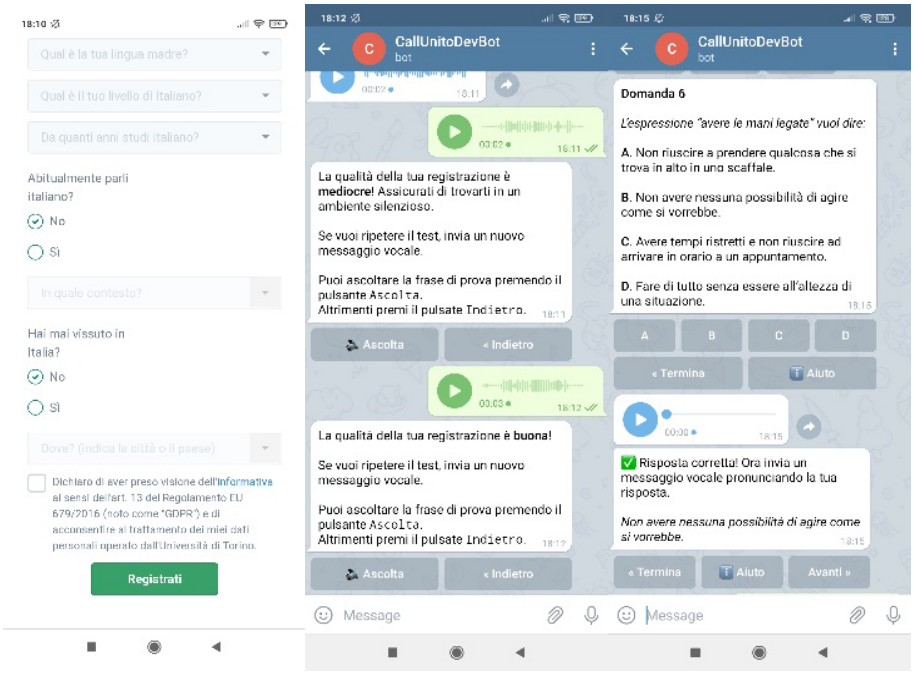
Due to the extemporaneous nature of the recorded speech signals and the automatic mode of the segmentation procedure operated by *Maus*, the vocalic segments detected for learners' utterances and those detected for the corresponding utterances of native speakers might differ. Therefore, before calculating the correlation, the segments of the two speakers are aligned on the basis of both the phonetic information they contain and their temporal position (Fig. 2). The correlation measure is then restricted to all and only those vowel segments to ensure phonetic-segmental homogeneity between the two speakers.

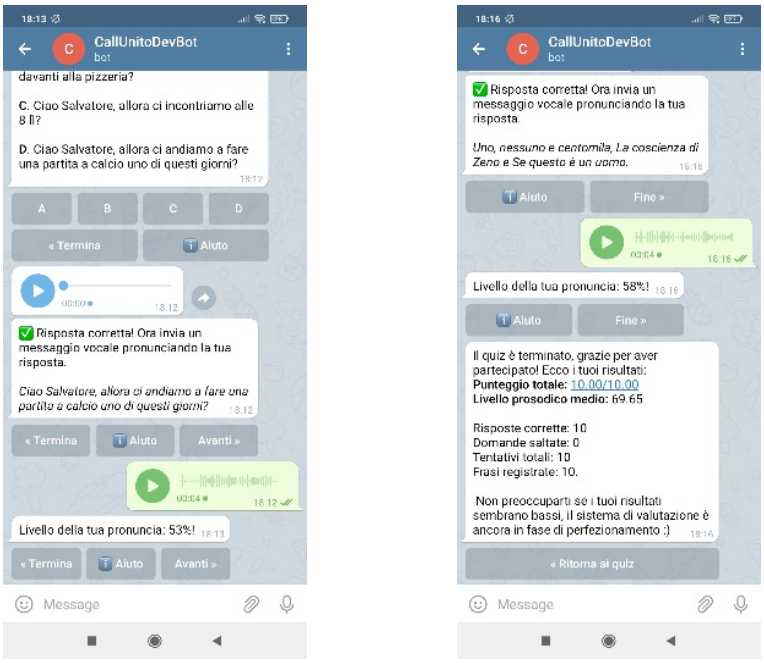
Figure 2 - *Segmented audio, temporal alignment and segment matching*



The administration of the task (Fig. 3) foresees a first registration phase in which the user indicates some of his sociolinguistic data (the linguistic level according to the CEFR, if and where he has lived in Italy, in which contexts he is used to speak Italian). Subsequently, after a microphone test demonstrating sufficient audio quality, the user can begin the task: each question is presented in written and oral form while the choice of the answer is among four options (see Appendix for the complete list of sentences).

Figure 3 - *Screenshots showing the main steps of the interaction chatbot-user*





If the user provides the wrong answer, he can try again and, once the correct answer is chosen, the chatbot suggests to record the answer through an audio message. At this point of the interaction, the elicited sentence is compared with those in the database and, after identifying the one with the highest intonation closeness, the chatbot returns the correlation value in percentage form. At the end of the task, a summary score obtained from the average of the values for each answer is calculated.

3. *Collected data and some first results*

A comparison with data from native speakers has been undertaken for several years (Delmonte, 2009) and deals well with the new ways of assessing automatic speech recognition systems. In this first phase of the project, we created a set of 10 questions and answers. Numbers, dates and medium complex intonation structures[1] (listed in the appendix) were included in the answer, which represents the comparison sentence, in order to have utterances of medium-high reading difficulty. The set was then read by 250 people who recorded question and answer using a mobile phone device leading to the creation of a database of 2500 sentences. The audio was then resampled to 16 kHz and, through manual segmentation, only the answers were segmented and labelled. The recorded samples came from different parts of Italy (half from Piedmont) and include 164 females and 84 males aged between 15

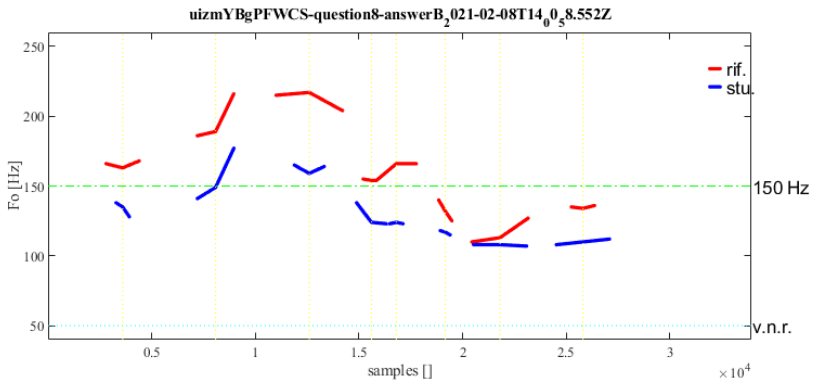
[1] This refers to a sentence with more than one intonation units (declarative, interrogative, continuative).

and 70 years[2]. The comparison allowed to find the utterance within the reference database that is closest to the one pronounced by the student and to give a correlation percentage[3]. It is important to underline that the percentage returned at the end of each given answer expresses exclusively the acoustic parameter responsible for the intonation, i.e. the fundamental frequency (f_0). Consequently, in order to offer a more accurate evaluation, the reference sentences produced by Italian speakers are evaluated by human assessors (Franco, Bratt, Rossier, Rao Gadde, Shriberg, Abrash & Precoda, 2010) in order to then correlate a selection of acoustic indices (such as speech rate, fluency or pausal scan) with the perceptual evaluation. Specifically, for each sentence we reported:

1. reading speed (slow, normal or fast);
2. regional accent (score from 1 -inconsistent accent- to 3 -marked accent);
3. word scansion (slow, normal or fast);
4. intonational fluency (scored from 1 -min spontaneity- to 3 -max spontaneity);
5. the age of the speaker (child, youth, adult or elderly).

The task was then submitted to 55 university students with heterogeneous linguistic profiles[4] (see Appendix for more detailed information). The average intonational proximity score is 72.9% with a minimum of 50% and a maximum of 86%[5], while the answer with the higher score is the sixth (74%). The graphical comparison of two intonation curves between the user and the reference sample (Fig. 4) with a correlation percentage of 94% shows, for example, an overall similar melodic pattern with some differences in the final part of the sentence.

Figure 4 - *Comparison of intonation curves between reference speaker (red) and user (blue) of the utterance "1/6 + 3/2"*



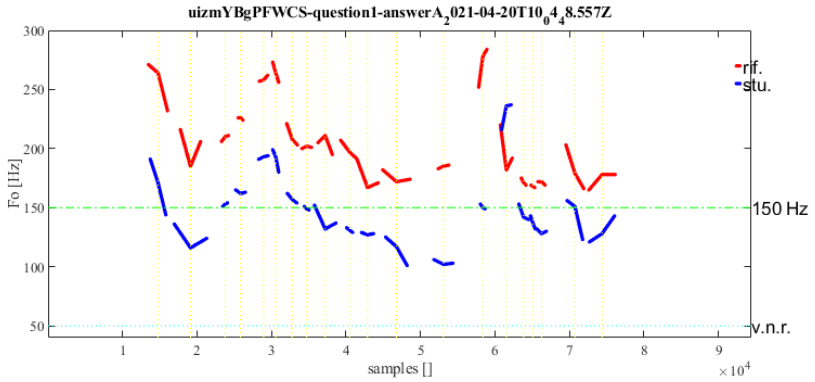
[2] This first phase of research allowed us to test the chatbot without considering specific variations related to the prosodic patterns: in terms of regional variation the corpus is still not balanced, but, on the other hand, the fact of belonging to a specific area did not necessary implicated a specific prosodic pattern (beside, the text was read).

[3] Where the student answered several times, only the last answer was considered.

[4] On the basis of the CEFR, one student reported having a A1 level, 9 students a B1, 15 students a B2 and the majority (26 students) a C1; 4 students did not give this information.

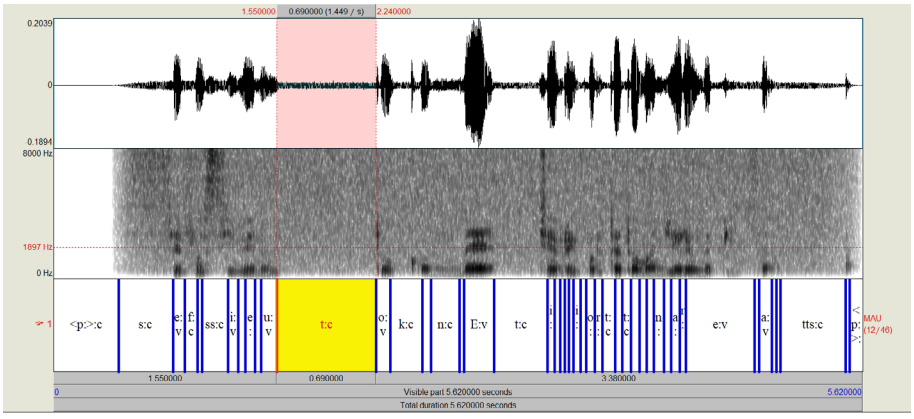
[5] The total sentences are not 550 but 532 because some students did not answer all the questions.

Figure 5 - *Comparison of intonation curves between reference speaker (red) and user (blue) of the utterance "Scusi, ho lasciato l'orologio a casa, sa dirmi che ore sono?"*



Even where the fundamental frequency is different (Fig. 5), the correlation allows us to evaluate speakers of different sex (in this case the score is 88%). Other factors such as hesitation within words (as in the example in Fig. 6) may come into play: this kind of phenomenon should be considered for a more coherent overall evaluation of the utterance.

Figure 6 - Praat *screen (wav and Textgrid) with the phonetic segmentation performed by* WebMaus *for the utterance "Se fossi venuto con me, ti avrei portato a mangiare una pizza."*



In a second step, we extracted some specific speech cues (see Tab. 1), in particular:

- number of syllables uttered by the speaker out of the duration of each speech-chain in the sentence (syllable rate, *SR*)
- number of syllables uttered by the speaker out of the total duration of the sentence (fluency rate, *FR*)
- total duration of pauses in seconds (*p. dur.*)
- number of pauses identified by the automatic segmentation procedure (*p.*)
- number of segments extracted by the automatic segmentation procedure (*seg.*)
- number of syllables extracted by the automatic syllabification procedure (*syl.*)

– number of segments extracted corresponding to vowel phonemes (*vow.*)
– duration of the utterance in seconds (*dur.*)

The sentences produced by native speakers vary from 2.5 to 4.8 ss while students take on average 1.3 ss longer than native speakers, up to a maximum of 2 ss longer in the first sentence. This sentence is also the one with the highest mean segment gap between native speakers and students (2.9) where on average students tend to add a maximum of 1.7 segments per sentence. The native speaker-student ratio of the number of syllables is between -0.1 (sentence 10) and 0.9 (sentence 1 and 8), with an increase in the number of vowels compared to the number of syllables for the first five answers. Speech rate is generally lower in students who also tend to take more pauses within the sentence (the sentence with the highest number of pauses is always the first one, with an increase of 1.1ss compared to native speakers).

Table 1 - *Extracted mean values of fluency rate, speech rate, pause duration, n° of pauses and duration of the 10 sentences grouped by reference speakers (blue) and students (orange)*

	1	2	3	4	5	6	7	8	9	10
FR	5.5	6.2	6.3	6.1	7.4	6.4	7	6.2	6.5	4.5
FR	4.4	5.4	5.6	5.4	5.9	4.9	6.1	5.7	5.4	3.4
diff.	1.1	0.9	0.7	0.7	1.4	1.5	0.9	0.6	1.1	1.1
SR	5.7	6.4	6.5	6.5	7.4	6.4	7	6.4	6.6	4.6
SR	4.9	5.8	5.9	5.9	6.1	5.1	6.3	5.9	5.5	3.7
diff.	0.8	0.7	0.6	0.6	1.3	1.4	0.7	0.5	1	0.9
p. dur.	0.5	0.5	0.4	0.6	0.3	0.3	0.3	0.4	0.3	0.3
p. dur.	1.6	1.1	1.2	1.3	0.9	0.9	0.9	0.9	0.9	1.2
diff.	1.1	0.6	0.8	0.7	0.6	0.6	0.6	0.5	0.6	0.9
p.	3.1	2.8	2.7	3.1	2	2	2	2.7	2.1	2.2
p.	4.6	3.3	3.3	3.7	2.2	2.2	2.4	2.7	2.3	2.9
diff.	1.4	0.6	0.6	0.6	0.2	0.2	0.4	0	0.2	0.7
dur.	4.7	3.7	4.6	4.8	3.8	3.1	3.5	3.6	3	2.5
dur.	6.7	4.9	5.9	6.1	5.3	4.6	4.5	4.6	4.2	3.5
diff.	2	1.2	1.3	1.3	1.4	1.5	1	1	1.2	1

Table 2 - *Extracted mean values of segments, syllables and vowels in the 10 sentences grouped by reference speakers (blue) and students (orange)*

seg.	51.9	42.5	60.7	59.3	59.7	41	44.8	45.5	40.3	19.1
seg.	54.8	43.8	62.4	60.8	61.3	41.2	45.4	46.5	40.5	19.6
diff.	2.9	1.3	1.7	1.5	1.6	0.2	0.6	1.1	0.2	0.5
syl.	23.6	21.1	26.5	27	26	17.9	21.8	20.5	17.8	8
syl.	24.5	21.6	27.2	27.6	26.1	18	22.1	21.3	17.9	7.9
diff.	0.9	0.5	0.7	0.6	0.1	0	0.2	0.9	0	-0.1
vow.	23.1	19.7	25.4	26.3	24.5	17.9	21.8	20.5	17.8	7.9
vow.	24.2	20.4	26.4	27.2	25.4	18	22.1	21.3	17.9	7.7
diff.	1.1	0.7	1	0.9	0.9	0	0.2	0.9	0	-0.2

4. *Discussion and future developments*

The assessment of prosody represents an important challenge in language teaching because it is intended to encourage the learner to reflect on how variation in acoustic indices in speech reflects changes in intended meaning. In this first phase of the project, we have focused on how to make a chatbot for language assessment effective and on what acoustic indices[6] to base it on in order to help the learner become aware of his or her prosodic competence in Italian. After describing its technical structure, we commented on some extracted values with the help of the first results of a preliminary study involving native Italian speakers and learners of Italian L2. Starting from the structure of the chatbot, we wanted to focus here on the presentation of the phases for the completion of the task. However, several steps of analysis remain that we plan to explore further in the next phase of the project. Having obtained reassuring results on a large sample, we are now in the phase of quantifying and solving the cases in which the utterances differ from the expected ones due to different word order and/or the presence of unintentional speech fragments (stuttering, hesitations, false starts) that cause differences such as to prevent the achievement of a minimum prosodic distance. Also, we are planning to implement a wider array of communicative functions in the chatbot which consider specific speech acts, types of focalization, effects of pauses. After completing the labelling of the database, we would like to focus on the sentences produced by the students for a phonetic-perceptual evaluation to be compared with the main acoustic indices at a later stage. In this regard, we are working on a more detailed feedback from the chatbot capable of returning specific parameters (speech rate, pauses) useful to the student. In a final step, we would like to increase and balance the reference corpus of native Italian speakers in order to expand the enunciative varieties in diatopic and diaphasic terms (Crocco, 2017). In this direction, a further step concerns the possible classification from which to train an algorithm capable of replacing the human operator along a line that distinguishes a more artefactual speech from a loose and spontaneous one (Nencioni, 1976; Voghera, 1989; Zmarich, Magno Caldognetto & Ferrero, 1996; Papi, Trentin, Gretter, Matassoni & Falavigna, 2020). Finally, an aspect that is still not very well studied concerns the reading of more complex utterances, which include the presence of simple mathematical formulae, acronyms and foreign words, without neglecting the expressions required by the statement pattern typical of some specialised languages. No specific training is provided for these aspects in the curricula for technical and scientific subjects, and no national publication provides exhaustive indications. For example, although everyone recognises graphic expressions such as "3/2", "12.1%", "011 6709718", "FBI", "report" etc., there is no publicly available source able to explain to a learner of Italian the most typical native speaker preferences for the pronunciation (or pronunciations) of these linguistic objects (Fry, 1989).

[6] We did not consider in this first step duration and intensity, which will be the object of a future work.

Acknowledgment

This project has received funding from Fondazione CRT – Bando Erogazioni Ordinarie 2020 – CALL-UniTO. We would like to thank all the participants who took part in the recording and the students who submitted to the task.

Bibliography

Bertinetto, P.M., Magno Caldognetto, E. (1993). Ritmo e intonazione. Introduzione all'italiano contemporaneo. In Sobrero, A.A. (Ed.), *Le strutture*, 2, Roma-Bari: Laterza, 141-192.

Boersma, P., Weenink, D. (2018). Praat: doing phonetics by computer [Computer program] Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/.

Busà, M.G. (2012). The role of prosody in pronunciation teaching: a growing appreciation. In Busà, M.G., Stella, A. (Eds.), *Methodological Perspectives on Second Language Prosody*, Padua: CLEUP, 101-105.

Canepari, L. (1983). *Italiano standard e pronuncia regionale*. Padova: CLEUP.

Canepari, L. (1985). *L'intonazione. Linguistica e paralinguistica*. Napoli: Liguori.

Cazade, A. (1999). De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues. In *Apprentissage des Langues et Systèmes d'Information et de Communication,* 2(2), 3-32. https://doi.org/10.4000/alsic.1623

Cheng, V, Lau, V., Lam, R., Zhan, T. & Chan, P. (2020). Improving English phoneme pronunciation with automatic speech recognition using voice chatbot. In Lee, L.U., Leong, H., Wang, F.L., Cheung, S.K.S., Au, O. & Li, K.C. (Eds.), *International Conference on Technology in Education*. Singapore: Springer, 88-99. 10.1007/978-981-33-4594-2_8

Chun, D. (1998). Signal analysis software for teaching discourse intonation. In *Language Learning & Technology*, 2, 6177.

Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.

Cresti, E. (1999). Force illocutoire, articulation topic/comment et contour prosodique en italien parlé. In *Faits de langue*, 13, 168-181.

Crocco, C. (2017). Everyone has an accent: standard Italian and regional pronunciation. In Cerruti, M., Crocco, C. & Marzo S. (Eds.), *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian*. Berlin, New York: Mouton de Gruyter, 89-117. 10.1515/9781614518839-004

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

DeBot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. In *Language and Speech*, 26(4), 331-350. https://doi.org/10.1177/002383098302600402

De Iacovo, V., Romano, A. (2019). Data-driven intonation teaching: an overview and new perspectives. In *EL.LE*, 8(2), 393-408. 10.30687/ELLE/2280-6792/2019/02/008

De Marco, A., Sorianello, P. & Mascherpa, E. (2014). L'acquisizione dei profili intonativi in apprendenti di italiano L2 attraverso un'unità di apprendimento in modalità

blended learning. De Meo, A., D'Agostino, M. (Eds.), *Varietà dei contesti di apprendimento linguistico, Studi AitLA* 1, 189-210.

De Meo, A., Pettorino M. (Eds.) (2012). *Prosodic and rhythmic aspects of L2 acquisition: The case of Italian*. Newcastle Upon Tyne: Cambridge Scholars Publishing.

Delattre, P. (1966). Les dix intonations de base du français. In *French Review*, 40, 1-14.

Delmonte, R. (2009). Prosodic tools for language learning. In *International Journal of Speech Technology*, 12(4), 161-184. 10.1007/s10772-010-9065-1

Fernoagă, V., Stelea, G.A., Gavrilă, C. & Sandu, F. (2018). Intelligent education assistant powered by chatbots. In *Proceedings of the 14th International Scientific Conference eLearning and Software for Education*, Bucharest, Romania, 19-20 April 2018, 2, 376-383. 10.12753/2066-026X-18-122

Franco, H., Bratt, H., Rossier, R., Rao Gadde, V., Shriberg, E., Abrash, V. & Precoda, K. (2010). Eduspeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. In *Language Testing*, 27(3), 401-418. https://doi.org/10.1177/0265532210364408

Frost, D., Picavet, F. (2014). Putting prosody first – Some practical solutions to a perennial problem: The Innovalangues Project. In *Research in Language*, 12(3), 233-243. https://doi.org/10.2478/rela-2014-0002

Fry, E.B. (1989). Reading formulas: maligned but valid. In *Journal of Reading*, 32(4), 292-297.

James, E.F. (1976). The acquisition of prosodic features of speech using a speech visualizer. In *International Review of Applied Linguistics in Language Teaching*, 14(3), 227-244. https://doi.org/10.1515/iral.1976.14.3.227

Kisler, T., Reichel, U. & Schiel, F. (2017). Multilingual processing of speech via web services. In *Computer Speech & Language*, 45, 326-347. https://doi.org/10.1016/j.csl.2017.01.005

Hirst, D. (1983). Structures and categories in prosodic representations. In Cutler, A., Ladd, R. (Eds.). *Prosody: models & measurement*. Berlin: Springer, 93-109.

Lacheret-Dujour, A. (2001). Modéliser l'intonation d'une langue. Où commence et où s'arrête l'autonomie du modèle? L'exemple du français parlé. In *Colloque international: Journées Prosodie 2001*, Grenoble, France, 10-11 Octobre 2001, 57-60.

Levy, M. (1997). *CALL: context and conceptualisation*. Oxford: Oxford University Press.

Lezhenin, Y., Lamtev, A., Dyachkov, V., Boitsova, E., Vylegzhanina, K. & Bogach, N. (2017). Study intonation: mobile environment for prosody teaching. In *IEEE International Conference on Cybernetics (CYBCONF)*, 3, 1-2. 10.1109/cybconf.2017.7985814

Martin, Ph. (2010). Learning the prosodic structure of a foreign language with a pitch visualizer. In *Speech Prosody 2010*, paper 980.

Moutinho de Castro, L., Coimbra, R., Rilliard, A. & Romano, A. (2011). Mesure de la variation prosodique diatopique en portugais européen. In *Estudios de fonética experimental*, 20, 33-55.

Munro, M.J. & Derwing, T.M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. In *Language Learning*, 49, 285-310. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nencioni, G. (1976). Parlato-parlato, parlato-scritto, parlato-recitato. In *Strumenti Critici*, X(1), 1-56.

Papi, S., Trentin, E., Gretter, R., Matassoni, M. & Falavigna, D. (2020). Mixtures of deep neural experts for automated speech scoring. In *Proceedings of Interspeech* 2020, 3845-3849. 10.21437/Interspeech.2020-1055

Pereira, J. (2016). Leveraging chatbots to improve self-guided learning through conversational quizzes. In *Proceedings of the fourth international conference on technological ecosystems for enhancing multiculturality*, TEEM '16, ACM Press, New York: 911-918. https://doi.org/10.1145/3012430.3012625

Romano, A., Giordano, G. (2017). Esperienze e riflessioni sulla didattica assistita dell'intonazione in italiano, inglese e francese. In Damascelli, A. (Ed.), *Digital resources, creativity, innovative methodologies and plurilingualism: new approaches to language teaching and learning*, Newcastle upon Tyne: Cambridge Scholars Publishing, 176-200.

Romano, A., Contini, M. & Lai, J.P. (2014). *L'Atlas Multimédia Prosodique de l'Espace Roman*: uno strumento per lo studio della variazione geoprosodica. In Tosques, F. (Ed.), *20 Jahre digitale Sprachgeographie*. Berlin: Humboldt-Universität-Institut für Romanistik, 27-51.

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In Ohala, J.J., Hasegawa, Y., Ohala, M., Granville, D. & Baile, A.C. (Eds.), *Proceedings of the International Conference of Phonetic Sciences* 1999, San Francisco, USA, 1-7 August 1999, 607-610.

Sorianello, P. (2006). *Prosodia: modelli e ricerca empirica*. Roma: Carocci.

Trouvain, J., Gut, U. (2007). *Non-native prosody. Phonetic description and teaching practice.* Berlin: Mouton de Gruyter.

Voghera, M. (1989). L'intonazione dell'italiano parlato spontaneo. In *The Italianist*, 9(1), 116-141. https://doi.org/10.1179/ita.1989.9.1.116

Zmarich, C., Magno Caldognetto, E. & Ferrero, F. (1996). Analisi confrontativa di parlato spontaneo e letto: fenomeni macroprosodici e indici di fluenza. In Cutugno, F. (Ed.), *Fonetica e fonologia degli stili dell'italiano parlato. Atti delle VII Giornate di Studio del "Gruppo di Fonetica Sperimentale" dell'Associazione Italiana di Acustica*. Roma: Esagrafica, 111-139.

Appendix

List of questions and answers:

1. Quale fra questi gruppi di opere è formato di sole opere italiane del Novecento?
   a. *L'Orlando furioso*, *L'Iliade* e *L'Odissea*.
   b. *La divina commedia*, *Il fu Mattia Pascal* e *Lo Zibaldone*.
   c. *I promessi sposi*, *La Vita Nuova* e *La Gerusalemme liberata*.
   d. *Uno, nessuno e centomila*, *La coscienza di Zeno* e *Se questo è un uomo*.

2. Chiara ha bisogno di sapere l'ora, come fa?
   a. Scusi, ho lasciato l'orologio a casa, sa dirmi che ore sono?
   b. Scusa, ho lasciato l'orologio a casa, per caso ha l'ora?
   c. Sa dirmi che ora sono per favore che ho lasciato l'orologio a casa?
   d. Mi scusi, ho lasciato l'orologio a casa, qual è il tempo oggi?

3. Marco e Salvatore si sono messi d'accordo per andare a mangiarsi una pizza e Marco deve dare appuntamento a Salvatore, cosa NON gli dirà?
   a. Ciao Salvatore, allora ci troviamo alle 8 lì davanti?
   b. Ciao Salvatore, allora ci becchiamo alle 8 davanti alla pizzeria?
   c. Ciao Salvatore, allora ci incontriamo alle 8 lì?
   d. Ciao Salvatore, allora ci andiamo a fare una partita a calcio uno di questi giorni?

4. Gianna vuole chiedere a Marta di comprare della frutta, cosa le dirà?
   a. Senti Marta, una domanda: ma a te piace la cassata?
   b. Senti Marta, stai uscendo? Se riesci a comprare della frutta, mi fai un favore.
   c. Senti Marta, mi sa che non ci sono più pesche.
   d. Senti Marta, che tu sappia, c'è ancora frutta in casa?

5. In quale anno l'Italia è diventata una Repubblica?
   a. Nel giugno del 1950.
   b. Durante l'autunno del 1939.
   c. Tra il 1945 e il 1946.
   d. Il 18 marzo 1861.

6. In quale giorno è scoppiata la Rivoluzione francese?
   a. Nell'anno 1789.
   b. Il 14 luglio 1789.
   c. Nel marzo 1938.
   d. Nel 1914.

7. L'espressione "avere le mani legate" vuol dire:
   a. Non riuscire a prendere qualcosa che si trova in alto in uno scaffale.
   b. Non avere nessuna possibilità di agire come si vorrebbe.
   c. Avere tempi ristretti e non riuscire ad arrivare in orario a un appuntamento.
   d. Fare di tutto senza essere all'altezza di una situazione.

8. In quale di queste frasi il periodo ipotetico è utilizzato correttamente?
   a. Se vieni con me, ti portassi a mangiare una pizza.
   b. Se venissi con me, ti avrei portato a mangiare una pizza.
   c. Se verresti con me, ti porto a mangiare una pizza.
   d. Se fossi venuto con me, ti avrei portato a mangiare una pizza.

9. Dov'è morto Giuseppe Garibaldi?
   a. Giuseppe Garibaldi è morto nell'isola di Caprera.
   b. Nel gennaio del 1882.
   c. Garibaldi è morto all'isola d'Elba.
   d. Quando aveva 75 anni.

10. 1 + 2/3 equivale a:
   a. 4/4 + 4/5
   b. 1/6 + 3/2
   c. ¼ + 4/9
   d. 1/3 – 0,75

Students' sociolinguistic profile[7]:

Gender	*Native language*	*Italian level*	*Years spent in Italy*	*Context*	*Abroad*	*Lived Where*
F	Spanish	C1	4	4	true	(PA)
F	Chinese	C1	4	4	true	(PI)
F	Spanish	C1	3	0	true	(PA)
F	Spanish	C1	3	0	true	(BO)
F	Spanish	B2	3		false	
M	Spanish	C1	2		true	(BG)
M	Spanish	C1	1	3	true	(PA)
F	French	B2	3	0	false	
F	French	C1	4	0	false	
F	French	B1	4	0	false	
F	French	C1	4	0	true	(MC)
F	French	B2	4		false	
F	French	B2	4	0	false	
F	French	B2	4		false	
F	French	C1	4		false	
F	French	C1	4		false	
M	French	B2	4	0	false	
M	French	C1	4	0	false	
M	French	B2	3	4	false	
F	French	C1	4	4	false	

[7] *Context* refers to the context where the student usually speaks Italian: 0 (university), 1 (private lessons), 2 (work), 3 (family), 4 (other), while *Abroad* means the student has spent some time in Italy (true) or not (false).

Gender	*Native language*	*Italian level*	*Years spent in Italy*	*Context*	*Abroad*	*Lived Where*
F	French	C1	4		false	
F	English	C1	3	3	true	(RM)
M	English	C1	4	4	false	
F	Spanish	B2	4		false	
F	French	C1	3	0	true	(TO)
F	Russian	B1	3	0	true	(RM)
F	Russian	B2	2		true	(RM)
F	French	B1	3	4	true	(RM)
F	French	B2	4	0	true	(RM)
F	English	B1	3		true	(TO)
F	Slovenian	B2	2	4	false	
F	Slovenian	B1	2		true	
F	Slovenian	C1	4	0	false	
O	Slovenian	B1	3		false	
M	English	B1	4	3	false	
F	Hungarian	C1	4	4	false	
F	English	B1	2	0	false	
M	Russian	C1	3		true	(VA)
F	English	A1	2	0	false	
M	English	B2	2	4	false	
F	Gaelic	B1	2		false	
M	English	C1	3	0	true	(BO)
F	Hungarian	C1	4	4	true	(TS)
M	Portuguese		null	0	true	(PI)
F	Spanish	B2	1	0,2,3	true	(AT)
F	Romanian		null	2,0	true	(TO)
F			null		true	(TO)
M	Armenian	B2	3		true	(TO)
F	Romanian	C1	4	2,3,0,4	true	(CN)
F	Portuguese	C1	4	2,3,0	true	(TO)
M	Romanian		null	2,3,4,0,1	true	(TO)
F	Serbian	C1	4	0,4	true	(TO)
M	Romanian	C1	4	3,0,4	true	(TO)
M	Portuguese	B2	1		true	(VC)
F	Spanish	C1	4	2,3,0	true	(TO)