

An acoustic-perceptual approach to the prosody of Chinese and native speakers of Italian based on yes/no questions

Marilisa Vitale^{1,2}, Philippe Boula de Mareuil², Anna De Meo¹

¹ Università degli Studi di Napoli “L’Orientale”, Naples, Italy

² LIMSI-CNRS, Orsay, France

vitalelem@unior.it, philippe.boula.de.mareuil@limsi.fr, ademeo@unior.it

Abstract

The present study investigates the prosody of yes/no questions (in comparison with statements) in Chinese learners and native speakers of Italian. Acoustic analyses and a perceptual test were performed, in order to identify the main trends in non-native productions. Results show the relevance of prosody, which differentiates elementary, intermediate and advanced Chinese learners of Italian. Listening tests based on prosody transplantation also suggest that non-native segments with a native Italian prosody are rated as less accented than are native Italian segments with a non-native prosody. Similar trends were found, overall, in terms of question/assertion discrimination, confirming the relative importance of prosody. These findings could be helpful for teachers and learners of Italian as a foreign language.

Index Terms: L2 Italian, yes-no questions, Chinese learners, prosody acquisition

1. Introduction

To acquire a good command of prosody in a second language (L2), it is crucial to identify the most salient patterns of the target language and to determine the main tendencies observed in the performance of L2 speakers. In L2 Italian, especially, the teaching of prosody often focuses on questions: this speech act is very important in everyday communication and it is introduced early in language classes [1].

A large body of research has concentrated on problems caused by question intonation. Ullakonoja [2] as well as Santiago-Vargas and Delais-Roussarie [3], for instance, were interested in the challenges posed by yes/no questions to Finnish learners of Russian and Mexican Spanish learners of French, respectively. The present study investigates problems faced by Chinese learners of Italian as an L2, when asking yes/no questions. It intends to better understand how speakers of a tone language perform in a non-tone language. Chinese questions have to preserve, to some extent, lexical tone pitch variations in order to keep the utterance meaning unchanged [4], whereas the Italian language distinguishes between questions and statements by intonation alone. Questions are usually marked by a rising-falling pitch movement on the last stressed syllable of the utterance (most often the penultimate syllable), at least in southern varieties of Italian [5] [6] [7]. Even though a common strategy is observed in the two languages under consideration in the present study, namely an overall higher pitch range in interrogative utterances [8] [9], interferences between the Italian and Chinese systems may affect the acquisition of prosody in the L2. For an overview on the prosody of yes/no questions, see [5] [7] for Italian, and [10] for Chinese.

This study compares statements and yes/no questions produced by both Chinese learners (of various proficiency

levels in Italian) and native speakers of Italian. It combines acoustic analyses and a perceptual test using prosody transplantation. This paradigm [11] consists of copying prosodic parameters from a native Italian utterance to a non-native one and vice versa. Unlike previously used techniques such as low-pass filtering [12], the method preserves comprehensibility: it thus yields a more ecological speech material. It was applied to various language pairings [13] [14]: in particular, it was used so as to disentangle the contribution of prosodic features (melody and global timing) and segmental features to the perception of accentedness and intelligibility in German-accented English and English-accented German. Here, the same methodology was used in order to evaluate the contribution of prosody to the perception of Chinese foreign accent and to the discrimination between questions and statements in Italian. The method, materials and results are described below.

2. Corpus

2.1. Speakers and material

For this study, 4 native Italian speakers and 12 Chinese learners of Italian (all females, aged 24 on average) were recorded. The learners, all speakers of Mandarin Chinese, belonged to three groups (each composed of 4 students) according to their level of competence in Italian (elementary, intermediate, advanced). All speakers lived in the Naples region (southern Italy).

Native and non-native speakers were recorded while reading a short dialogue in Italian, including 7 yes/no questions. These questions varied in length (between 5 and 11 syllables each), but they all ended with a paroxytone (that is, a word stressed on the penultimate syllable). This choice was determined by the need to avoid stress-related variation and by the frequency of this stress pattern in Italian: the Italian vocabulary is mostly composed of paroxytone words (in over 75% of cases [15]). Speakers were then instructed to read the same questions in the declarative modality — word order does not change in Italian. The corpus analysed below was therefore made up of 112 questions and 112 statements. Examples of sentences are given in § 3.2.

2.2. Acoustic analysis

The 224 resulting utterances (as many questions as statements) were analysed to identify prosodic differences between native and non-native speakers of Italian, as well as across the three learner groups. For each utterance, mean pitch, the number and duration of inter-pause speech intervals, the number of syllables per inter-pause speech interval, the duration of silent pauses and disfluencies were computed. The speech articulation rate (excluding pauses), phonation rate (including pauses), tonal range (between maximum and

minimum defined pitch values) and the percentage of disfluencies were calculated. Furthermore, fundamental frequency (f_0) values were measured at the midpoint of each of the last three syllables, for each utterance. These measurements were taken using the Praat software [16].

Mean pitch differences in semitones (ST) between questions and statements were averaged for each speaker group. In each group, questions are higher than statements: by 0.5 ST for elementary learners, 0.2 ST for intermediate learners, 1.7 ST for advanced learners and 0.8 ST for native speakers. Hence, no clear tendency seems to emerge.

Results of the other acoustic analyses are shown in Table 1. As expected, the speech articulation rate and phonation rate of elementary learners are lower than those of the other two learner groups and those of native speakers. Note that already at an intermediate level of language proficiency, Chinese learners are able to properly manage duration-related variations. Intermediate learners, together with the advanced learner group, show values very close to the native model, partly due to our native speakers' tendency to hyperarticulate.

Table 1. *Acoustic measurements: articulation rate (syll./s), phonation rate (syll./s), tonal range (semitones), % dis. (percent time of disfluencies) within questions (Q) and statements (S).*

Speaker group	Q	S
Articulation rate		
Elementary L2	3.8	3.8
Intermediate L2	5.2	5.1
Advanced L2	5.5	5.0
Native Italian	5.4	5.3
Phonation rate		
Elementary L2	3.3	3.7
Intermediate L2	5.1	5.0
Advanced L2	5.5	5.0
Native Italian	5.4	5.3
Tonal range		
Elementary L2	12.6	8.3
Intermediate L2	7.2	7.9
Advanced L2	7.7	7.2
Native Italian	9.4	13.7
% dis.		
Elementary L2	3.2	0.0
Intermediate L2	0.1	0.2
Advanced L2	0.0	0.0
Native Italian	0.0	0.0

As for the tonal range of both questions and statements, the three groups of Chinese learners seem to be unable to produce varied pitch movements as native Italian speakers do. On average, non-native speakers' tonal range is more reduced than that of native speakers — by 4 ST. Interestingly, the only exception stems from questions asked by elementary learners. Because of their linguistic insecurity (their uncertainty and lack of confidence in how to pronounce words properly), these speakers tend to produce creaky-voiced vowels and sometimes a high-rising terminal tune, resulting in a tonal range even wider than that of native speakers.

As far as disfluencies are concerned, they mainly come from elementary learners of Italian. Due to the use of read speech, they are very few (only 3% of utterance duration). They mainly consist of self-repairs and repetitions.

In addition to this, the final pitch contours of the 224 sentences of the corpus were analysed in detail. Results are displayed in Figure 1 in semitones calculated with respect to the minimum f_0 measure (172 Hz). The comparison of Fig. 1a and 1b shows that the involved native Italian speakers' terminal pitch range is higher in questions than in statements (with a 3.6 ST difference). The average f_0 range difference between questions and statements is 3 ST in the case of advanced learners. For the other two groups of Chinese speakers, f_0 values are quite similar between the speech acts considered: the measured increase in questions is only 1.1 ST for intermediate learners and 0.7 for elementary learners.

In question-final syllables, the involved native speakers produced a slight upward-downward curve, which is the typical pitch movement of the Italian language (or at least its Neapolitan variety), as evidenced by the literature (e.g. [7]). The group of elementary learners, on average, displayed no movement at all but instead produced a flat pitch curve. It should be noted however that, in contrast with this general trend, several beginners' questions exhibit a rising pitch movement on the very last syllable, which sounds foreign as will be confirmed by the perception test. A flat pitch pattern can also be found in the intermediate speakers' productions, whereas the advanced learners' pitch rise on the penultimate (stressed) syllable is closer to the native Italian model (see Figure 1a).

In statements, a sharp pitch fall is noticeable only in native Italian speakers. The three groups of Chinese learners, on average, show gradual downstep towards the utterance-final syllable (see Figure 1b).

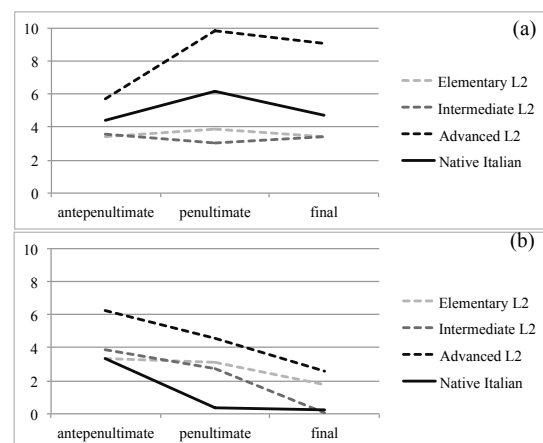


Figure 1: *Last three syllables (average values in semitones) of questions (a) and statements (b).*

3. Perceptual test

3.1. Protocol and listeners

A perceptual test was then conducted to check whether and to what extent there was an audible improvement of prosody from Chinese elementary, intermediate and advanced learners to native Italian speakers, in terms of native-likeness and communicative effectiveness. More precisely, this test aimed at evaluating speakers' degree of foreign accent and

possible difficulties encountered by Italian listeners in properly identifying questions and statements. These difficulties are expected to be primarily related to prosody, rather than to the segmental level. As regards the perception of a foreign accent, it may be due to both prosodic and segmental levels [11] [14]. For this purpose, the prosody transplantation technique was used so as to exchange native and non-native acoustic-prosodic features, phoneme by phoneme. It relied on the PSOLA algorithm implemented in Praat, which allows phoneme durations to be matched, between two realisations of a given utterance, before grafting the f_0 curve of version 1 onto version 2 and vice versa.

The perceptual test was administered to 40 native Italian listeners (18 males and 22 females, aged 34 on average) with no reported history of hearing impairment, coming from various regions of Italy. Almost all of them (34 out of 40) were experts in the field of language and phonetics research. However, none of them could speak an Asian or a tonal language.

3.2. Material and tasks

Out of the 84 questions produced by Chinese learners of Italian (7 × 4 per group), 12 questions were used for the perceptual test. Nine of these sentences, uttered by 3 different speakers from each of the 3 learner groups, had the following structures: *È una cosa che si usa spesso?* 'Is it something often used?', *Questo oggetto si trova in casa?* 'Is this object in the house?', *Si usa quando si è stanchi?* 'Is it used when somebody is tired?'. Another 3 questions, uttered by another speaker from each of the 3 learner groups, had the following structure: *È un(a)* 'Is it a' + Noun, with Noun being *divano* 'sofa', *persona* 'person' or *oggetto* 'object'. In addition, 12 questions were selected among the native Italian speakers' corresponding renditions, in such a way that each learner group was associated with 4 different Italian voices producing each of the 4 selected questions. Prosody transplantation was then performed, thus generating 24 additional stimuli. The same type of selection and prosody transplantation procedure was carried out for statements. The question + statement samples finally used in the perceptual test were therefore composed of:

- 12 Q + 12 S original non-native Italian utterances (4 + 4 elementary, 4 + 4 intermediate and 4 + 4 advanced);
- 12 Q + 12 S original native Italian utterances, selected so as to never have the same voice repeating the same utterance and to have different voices associated with each of the three learner groups;
- 12 Q + 12 S utterances with native Italian segments and a non-native prosody;
- 12 Q + 12 S utterances with non-native segments and a native Italian prosody.

The 96 resulting stimuli were administered to the 40 listeners in different random orders, through an online interface (http://www.audiosurf.org/test_perceptif_marilisa/). Participants were informed that the experiment dealt with the Italian language spoken by native and non-native subjects, and that they would listen to excerpts of original or acoustically-modified speech. They were advised to use headphones or earphones.

Subjects first provided autobiographical information (age, education, place of residence, etc.). Also, they were asked very

general questions, before a short familiarisation phase with the types of stimuli. They first listened to examples of native/non-native, original/manipulated statements and questions (not used in the actual test). For each utterance, they were then asked to:

- assess the degree of foreign accent on a continuous 0–5 scale (0 = no foreign accent; 5 = very strong foreign accent);
- identify the correct speech act, discriminating between "question" and "statement".

To accomplish the first task, a slider was provided (by default located in the middle of the scale); for the second task, participants had to click on buttons. They could listen to each stimulus as many times as they needed, but it was not possible to correct previous answers once a new stimulus was displayed.

At the end of the test, listeners were asked other questions (1) to indicate the speakers' most salient and characteristic linguistic cues and (2) to identify the native and non-native speakers' geographical and linguistic backgrounds.

3.3. Results

The results of the perceptual test are reported in Table 2. On the basis of original stimuli, analyses of variance (ANOVAs) were carried out separately on listeners' responses, in terms of rating and correct speech act identification, with the two fixed factors Modality (question or statement) and Level group (elementary, intermediate, advanced or native). On the basis of prosody-transplanted stimuli, a second series of ANOVAs was performed on listeners' responses, with the two fixed factors Modality (question or statement) and Type of stimulus (6 levels: see Table 2b).

The degrees of foreign accent attributed by listeners to the original stimuli gradually decrease from elementary learners to native speakers of Italian (Table 2a). The Italian pronunciation of Chinese speakers improves linearly, in both questions and statements, and L1 Italian speakers (with a 0.1 degree of foreign accent) are properly identified as native speakers. Differences related to Modality are not significant, but differences across Level groups are significant [$F(3, 1960) = 2387; p < 0.001$] — the interaction between the two factors is marginal. Tukey's HSD post hoc test shows that all differences across Level groups are highly significant.

Responses in terms of question/statement discrimination display slightly different patterns, with significant differences related to Modality [$F(1, 3928) = 10; p < 0.001$], significant differences across Level groups [$F(3, 3928) = 123; p < 0.001$] and a significant interaction between the two [$F(3, 3928) = 12.7; p < 0.001$]. As in accent ratings, a linear progression is observed in the speech act identification of statements. In contrast, beginners' questions are better recognised than are intermediate students' questions, probably due to the fact that the elementary level speaker group sometimes produces high-rising terminal tunes. These pitch rises on utterance-final syllable rather than on the last stressed syllable of the question do sound foreign but they may be effective from a communicative point of view.

The use of acoustically modified stimuli allowed us to tease apart the influence of prosodic (suprasegmental) level and segment articulation in terms of both foreign accent rating and question/statement discrimination. As shown in Table 2b,

the stimuli with non-native segments and a native Italian prosody are perceived as having a slight foreign accent, whereas the stimuli with native Italian segments and a non-native prosody are perceived as more strongly foreign accented. Differences related to Modality are not significant, but the Type of stimulus has a major effect [$F(5, 1908) = 101$; $p < 0.001$] — the interaction between the two being marginal. Post hoc comparisons (Tukey's HSD test) show that stimuli containing native Italian prosody form a homogeneous subset, but differ significantly from the other stimuli.

A similar trend is observed concerning the speech act identification of questions vs. statements. Differences related to Modality are still not significant, but the Type of stimulus has a major effect [$F(5, 1908) = 87.1$; $p < 0.001$] and the interaction between the two is here significant [$F(5, 1908) = 2.23$; $p < 0.05$]. Subsequent post hoc comparisons (Tukey contrasts) show that stimuli with non-native segments and a native Italian prosody are discriminated significantly better than stimuli with native Italian segments and a non-native prosody. Both questions and statements with the native Italian prosody turn out to be identified almost perfectly by the listeners, regardless of possible mispronunciations at the segmental level. By contrast, the superimposition of the non-native rhythm and intonation on native Italian productions dramatically reduced question/statement discrimination.

Table 2. Average degree of perceived foreign accent and speech act identification of the original (a) and manipulated (b) stimuli Q = questions; S = statements.

(a) Original stimuli		
Speaker group	Q	S
	Degree (/5)	
Elementary L2	4.3	4.1
Intermediate L2	3.7	3.6
Advanced L2	2.6	2.6
Native Italian	0.1	0.1
	% Q/S Id	
Elementary L2	59	48
Intermediate L2	41	72
Advanced L2	73	82
Native Italian	100	99

(b) Manipulated stimuli			
Prosody donor	Prosody receiver	Q	S
		Degree (/5)	
Elementary L2	Native Italian	3.6	3.5
Intermediate L2	Native Italian	2.4	2.7
Advanced L2	Native Italian	1.8	1.8
Native Italian	Elementary L2	1.6	1.4
Native Italian	Intermediate L2	1.8	1.6
Native Italian	Advanced L2	1.7	1.5
Prosody donor	Prosody receiver	Q	S
		% Q/S Id	
Elementary L2	Native Italian	59	56
Intermediate L2	Native Italian	62	73
Advanced L2	Native Italian	69	76
Native Italian	Elementary L2	98	98
Native Italian	Intermediate L2	97	98
Native Italian	Advanced L2	99	96

In their final comments, most listeners (about 30) mentioned intonation and rhythm among the features which helped them make their decisions. Also, 30 subjects properly identified the regional origin of native speakers (i.e. Campania) and the linguistic background of non-native speakers (i.e. Chinese). Most of them reported an average familiarity with Chinese-accented Italian, which suggests that their evaluations were relevant.

4. Conclusions

This study based on Chinese learners of Italian as an L2 included acoustic analyses and a perception test in which suprasegmentals took precedence over segmentals, as in a previous study focusing on Spanish-accented Italian [11]. It highlighted the importance of focusing on prosody in order to improve both native-likeness and communicative effectiveness, at least as far as the perception of questions and statements is concerned. Overall, the more proficient the speakers, the better their questions are judged. However, results suggest that Chinese learners of Italian, a non-tonal language, succeed in producing yes/no questions in an appropriate manner only if they have acquired an advanced level of proficiency in the L2. It therefore seems important to point out the main problems faced by Chinese speakers of Italian, in order to give them feedback when learning the Italian yes/no question prosody.

Primarily, the pitch range of questions and statements need to be differentiated, since it is usually more extended in the case of questions. When asking questions ending with paroxytone words (a frequent pattern in Italian), Chinese learners of Italian should also be taught to realise a rising-falling pitch movement on the penultimate, prominent syllable of the utterance. Particular attention must be paid to beginners' tendency to produce a pitch rise on the utterance-final syllable: even if this intonational movement is communicatively effective in the sense that it is prone to be interpreted as a question contour, it is far from the native Italian model and sounds foreign.

The Chinese speakers under investigation here were also recorded in their L1. Acoustic analyses of their native productions are currently in progress and provide interesting comparisons. Finally, the results presented here need to be validated by further studies on spontaneous speech, to provide a clearer picture of question intonation in Italian as a foreign language. The understanding of questions in conversational speech will undoubtedly be another challenge.

5. Acknowledgements

This work was done during the first author's stay at LIMSI-CNRS, financed by an Erasmus Placement grant. We are grateful to all the speakers and listeners who volunteered to take part in the experiments.

6. References

- [1] Spinelli, B. and Parizzi, F., *Profilo della lingua italiana. Livelli di riferimento del QCER A1, A2, B1, B2*, La Nuova Italia, Florence, 2010.
- [2] Ullakonoja, R., "How do native speakers of Russian evaluate yes/no questions produced by Finnish L2 learners?", *Rice Working Papers in Linguistics*, 94(2):92–105, 2010.
- [3] Santiago-Vargas, F. and Delais-Roussarie, É., "The acquisition of question intonation by Mexican Spanish learners of French", in É. Delais-Roussarie, M. Avanzi, S. Herment [Eds], *Prosody and languages in contact. L2 acquisition, attrition, languages in multilingual situations*, Springer Verlag, Berlin, 2014 (to appear).
- [4] Visceglia, T. and Fodor, J. D., "Fundamental frequency in Mandarin and English: Comparing first- and second-language speakers", in C. Lleao [Ed.], *Interfaces in Multilingualism: Acquisition and Representation*, John Benjamins, Amsterdam, pp. 27–59, 2006.
- [5] D'Imperio, M., "Italian intonation: An overview and some questions", *Probus*, 14:37–49, 2002.
- [6] Grice, M., D'Imperio, M., Savino, M., Avesani, C., "Strategies for intonation labelling across varieties of Italian", in S.-A. Jun, [Ed.], *Prosodic typology: the phonology of intonation and phrasing*, Oxford University Press, Oxford, pp. 55–83, 2005.
- [7] Savino, M., "The intonation of polar questions in Italian: Where is the rise?", *Journal of the International Phonetic Association*, 42(1):23–48, 2012.
- [8] Bolinger, D., "Intonation across languages", in J. Greenberg [Ed.], *Universals of human language (Volume 2: Phonology)*, Stanford University Press, Stanford, pp. 471–524, 1978.
- [9] Gussenhoven, C., "Intonation and interpretation: phonetics and phonology", in *Proceedings of the 1st International Conference on Speech Prosody*, pp. 47–57, 2002.
- [10] Yuan, J., "Mechanism of Question Intonation in Mandarin", in Q. Huo, B. Ma, E.-S. Chng, H. Li [Eds], *Chinese spoken language processing*, Springer Verlag, Berlin, pp. 19–30, 2006.
- [11] Boula de Mareüil, P. and Vieru-Dimulescu, B., "The contribution of prosody to the perception of foreign accent", *Phonetica*, 63(4):247–267, 2006.
- [12] Munro, M. J., "Nonsegmental factors in foreign accent: Ratings of filtered speech". *Studies in Second Language Acquisition*, 17:17–34, 1995.
- [13] Holm, S., "Intonational and durational contributions to the perception of foreign-accented Norwegian: An experimental phonetic investigation", PhD thesis, Norwegian University of Science and Technology, Trondheim.
- [14] Winters, S. and O'Brien, M.G., "Perceived accentedness and intelligibility: The relative contributions of F0 and duration", *Speech Communication*, 55(3):486–507, 2013.
- [15] D'Imperio, M. and Rosenthal, S., "Phonetics and phonology of main stress in Italian", *Phonology* 16(1):1–28, 1999.
- [16] Boersma, P. and Weenink, D., "Praat: doinghonetics by computer" [Computer program], Version 5.3.15 retrieved 22 May 2012 from <http://www.praat.org/>, 2012.