The interplay of tone and intonation: $f_0$ contours produced by Hungarian learners of Mandarin

Kornélia Juhász, Huba Bartos

Hungarian Research Centre for Linguistics
Eötvös Loránd University, Hungary

juhasz.kornelia@nytud.hu, bartos@nytud.hu

Abstract

In this acoustic analysis we compare the $f_0$ contours of syntactically unmarked Mandarin Chinese (MC) echo-questions to statements in the production of two groups of Hungarian learners of Chinese and a native MC control group. MC and Hungarian differ in terms of question intonation patterns, thus we hypothesized that synchronizing tone and intonation production poses problems to L2 learners of MC due to the different intonation patterns and the absence of tones in the L1. In the experiment speakers were asked to contrast the same 4-syllable utterance in the two analysed moods: as a declarative and as a syntactically unmarked echo-question. We compared $f_0$ contours of each isolated syllable by general additive mixed models. Our results partially confirmed the hypothesis that L2 learners’ production MC questions are not as elevated in their characteristic as natives, yet L2 speakers do differentiate between the two moods in their production but produce terminal tonal patterns differently than those of natives.

Keywords: L2 production, Mandarin Chinese prosody, tones & intonation interaction, unmarked Mandarin Chinese questions

1. Introduction

In this acoustic analysis, we aim to shed light on how Hungarian learners of Mandarin Chinese contrast the intonation patterns (conceived as $f_0$-contours) of unmarked yes-no questions and statements, by analysing 4-syllable utterances having the same tonal value, either rising (T2) or falling (T4), throughout. In tonal languages, such as Mandarin Chinese (MC), $f_0$ serves for the realization of both lexical tones and intonation [1]. This means that, on one hand, $f_0$ modulation is locally dependent primarily on tone values (or tonal contexts), further affected by segmental effects, as well. On the other hand, these local effects also interact with intonation patterns, yielding the actual $f_0$ contour [2]. Finally, adjacent tones are also closely dependent on each other (tonal coarticulation).

Taking a broader view, in distinguishing MC statement and question intonation patterns, localized (e.g., terminal rise on the last (tonal) syllable) and global acoustic cues (raised $f_0$ over the whole utterance) have been identified concerning $f_0$ register and $f_0$ range [3]. Regarding $f_0$ register, according to Shen’s MC intonation model [1], statements display a gradually descending pattern, while unmarked yes-no questions feature a significantly higher $f_0$ throughout the whole utterance (compared to the declarative contour), complemented by a terminal rise. According to Shen’s study, this elevated characteristic of the question intonation pattern affects not just the top line of the contour (the peaks of each syllable strung together), but also the base line (the valleys of each syllable), thus there is no expansion of the $f_0$ range. The aforementioned terminal rise can be attributed to the lack of lexical/syntactic cues (such as particle ma), thus prosodic cues are used exclusively to express interrogative force [3, 4, 5]. However the realization of the terminal rise is tone-dependent, meaning that in case of T2, the tonal $f_0$ curve is realized with a widened pitch range, which rises much higher than in statements. In the case of T4 the $f_0$ range remains intact both in question and statement intonation, but the $f_0$ contour is realized at a higher $f_0$ level for questions [4, 5].

In contrast to MC, Hungarian is a non-tonal language, thus $f_0$-change is manifest only at clause level. In addition to this, while in Hungarian the declarative pattern is realized in a descending manner similar to MC, the prosodic structure of $f_0$ contour in yes/no questions differs in MC and Hungarian: in the latter, the initial $f_0$ value is low (compared to declaratives), and the $f_0$ contour is characterized by a rising structure peaking on the penultimate syllable (unless fewer than two syllables follow the final phrasal stress), followed by a fall [6, 7, 8] (Fig. 1.).
2. Hypotheses & Method

2.1. Hypotheses

In L2 acquisition, prosodic patterns, e.g. intonation patterns, are transferred from L1 [9, 10]. Due to L1 prosodic transfer, on the basis of the different structures of intonation patterns and the absence of tones in L1, we expect that synchronizing tone and intonation production poses problems to Hungarian learners. In particular, we hypothesize that Hungarian learners of MC favor the proper tone-production over intonation, in this manner producing unmarked questions similar to statements, without an elevated $f_0$. Besides, we are seeking an answer to the question how language learners’ $f_0$ contours are shaped in T2 and T4 sequences.

2.2. Method

We have analysed three adult speaker groups (5 female speakers per group, altogether 15 speakers): 1. Hungarians with cca. one year language experience of MC, second year Chinese undergraduates (beginners); 2. Hungarians with 3-4 years of learning MC, attending a Chinese Studies Masters programme (advanced learners), and 3. a control group of Chinese natives (from within a 300 km range around Beijing). Regarding the material, we have compared the production of declarative (hereinafter: D) and syntactically unmarked yes/no interrogative broad focus sentences (echo-questions, hereinafter: Q), all SVO with 4 syllables (2-1-1 syllables), matching pairwise in syllable structure and the number of voiced segments. Each syllable in a sentence has the same tonal value (either T2 or T4). Three interrogative and three declarative sentences for each tone were read for three times, presented as short question-answer dialogues, projected on a screen with Chinese characters and pinyin. The sentences were introduced to the speakers before recordings were made. Also, the instructions emphasized that the aim of the experiment was contrasting question and declarative intonation. We recorded a total of 3600 syllables (2 tones × 3 sentences × 5 repetitions × 2 modalities × 15 speakers).

Table 1: The echo-questions (UMQ) and declaratives (D) that were read in short question-answer dialogues.

<table>
<thead>
<tr>
<th>Q&amp;D</th>
<th>Time 2</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>Wang jue has not a boat.</td>
<td>Lang Que sells bracelets.</td>
</tr>
<tr>
<td></td>
<td>Does Wang jue have a boat?</td>
<td>Does Lang Que sell bracelets?</td>
</tr>
<tr>
<td>UMQ</td>
<td>Wang jue has not a house.</td>
<td>Lang Que sells electricity.</td>
</tr>
<tr>
<td></td>
<td>Does Wang jue have a house?</td>
<td>Does Lang Que sell electricity?</td>
</tr>
<tr>
<td>D</td>
<td>Wang jue has not a land.</td>
<td>Lang Que sells tickets.</td>
</tr>
<tr>
<td></td>
<td>Does Wang jue have a land?</td>
<td>Does Lang Que sell tickets?</td>
</tr>
<tr>
<td>UMQ</td>
<td>Wang jue has not a spring?</td>
<td>Does Lang Que sell tickets?</td>
</tr>
<tr>
<td></td>
<td>Does Wang jue have a spring?</td>
<td>Does Lang Que sell tickets?</td>
</tr>
<tr>
<td>D</td>
<td>Wang jue does not have a spring.</td>
<td>Lang Que sells tickets.</td>
</tr>
</tbody>
</table>

Within voiced segments, $f_0$ was extracted by 5 millisecond intervals automatically in Praat [11]; since the the onset of the 2nd and 4th syllables was a voiceless segment, it was excluded from the analysis. The extracted $f_0$ values were converted to semitones (with a reference value of 50 Hz [12], with hqmisc package [13] in R [14]) and $f_0$ curves were analysed by generalized additive mixed models (GAMM [15]) also in R [14], using the packages mgcv [16] and itsadug [17]. Matching syllables of questions and answers were analysed together. For each syllable pair (equal to 8 models), a basic model was built where the $f_0$ (in semitones) is the dependent variable and the normalized duration was the independent variable, and the models were further complemented by a combined parametric factor variable of the modality produced by different speaker groups (equal to 6 levels), and a random smooth function was applied to each $f_0$-trajectory, respectively. The combined factor variable was ordered in the following way: In the case of all four syllables we chose the native declarative curve (D_Nat) as the reference, and difference curves were computed compared to its realization, hence the ordered levels of this factor variable are: native D, native Q, advanced D, advanced Q, beginner D, beginner Q. Regarding the parametric coefficients in the Results section, the overall estimated mean is shown by each level of the factor variable (so basically the average height of the reference curve (D_Nat) and the difference curves computed from the reference) with $t$-tests’ significance levels. (In the Results section for the sake of practicality, we present the actual estimated means for each factor level, and not their difference from the reference, which is the original output of the model summary). The approximate significance of the smooth terms indexes whether the shape of the $f_0$ curve differ from the reference curve. The models were treated for autocorrelation [17]. Moreover, each model was checked by the gam.check() command [16], and the residuals were normally distributed, and even if the $k$-index was above 1, the $p$-values were not significant in either case.
3. Results

3.1. T2 sequences

Regarding the results of T2 sequences, we can observe a scissor-like opening between Qs and Ds towards the end of the utterance not just in the curve shapes, but in the case of the estimated means as well, where the difference between the two moods are only less than 1 semitone in the first two syllables, however reaches 5 semitones on the last syllable (Table 2, Fig. 2.). Concerning the overall utterance structure trends by considering the estimated means in Table 2., native speakers’ feature an overall descending structure for both Qs and Ds, however terminal rise appears on the last syllable of Qs, due to the T2’s expanded range (Fig. 2.). Regarding L2 learners’ production in overall trends, beginners approximate native production, in a manner that descending pattern is observed for both modalities, and the last two syllables of Qs and Ds yield with an average difference of 3 semitones. In contrast, although advanced learners do produce terminal rise as well on Qs (compared to Ds), however the f0-difference between the two moods is only less than 2 semitones, still, the largest difference between the two moods is positioned to the last syllable. As regards to the smooth terms, all f0 curve differ significantly from native declarative curve (Table 2, Fig. 2.), from which we can conclude that neither group produce neither syllable’s declarative curve as natives. Moreover, regarding the comparison between the two modalities within groups, we found significant differences between the two moods, where at least 80% of Qs’ curves are realized in a significantly higher f0-frequency compared to Ds, (with one exception: the 1st syllable of natives, where the first 42% of the Q and D curve is not distinguished). More particularly, while natives produce at least 2 semitone-difference between the Q and D curves, L2 learners rather elevate Qs with only 1 semitone higher compared to Ds. Additionally, the two L2 learning groups follow overall different patterns contrasting Qs and Ds: while advanced learners stick to almost the same shape in Qs and Ds, beginners’ production varies inconsistently between the two moods (Fig. 2.). Comparing each syllable in the production of both L2 learner group, the rising phase is apparent on the 1st, however is lacking on both the 2nd and 3rd syllable, causing the initiation of the terminal rise to delay, and exclusively realize on the ultimate syllable (as opposed to the native production, where the penultimate syllable is also involved in the terminal rise). As regards to the ultimate syllables’ f0 patterns, beginners approximate more natives’ production than advanced learners, with a steep linear, and less concave rise. In contrast, advanced learners’ production does not feature an apparent rising phase terminally. Concerning the effectiveness of the models signalled by the $R^2$ value, in each case, more than 90% of the dependent variables’ variance was explained by the analysed independent and random effects.

Table 2. The parametric coefficients (the estimated mean (Est.), the t-value and the significance (above) and the approximate significance of smooth terms (the effective degrees of freedom (edf) and the F- and p-values) for the four T2 syllable, and $R^2$ for all models (below)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>t</td>
<td>Sig.</td>
<td>Est.</td>
</tr>
<tr>
<td>D_Nat</td>
<td>24.4</td>
<td>94.5</td>
<td>***</td>
<td>24.0</td>
</tr>
<tr>
<td>Q_Nat</td>
<td>23.2</td>
<td>6.9</td>
<td>***</td>
<td>24.9</td>
</tr>
<tr>
<td>D_Adv</td>
<td>24.8</td>
<td>0.9</td>
<td>n.s.</td>
<td>23.9</td>
</tr>
<tr>
<td>Q_Adv</td>
<td>25.8</td>
<td>3.6</td>
<td>***</td>
<td>24.4</td>
</tr>
<tr>
<td>D_Beg</td>
<td>24.4</td>
<td>0.1</td>
<td>n.s.</td>
<td>23.9</td>
</tr>
<tr>
<td>Q_Beg</td>
<td>25.5</td>
<td>2.9</td>
<td>***</td>
<td>24.4</td>
</tr>
</tbody>
</table>

**Figure 2:** The estimated question (red solid curve) and declarative (blue dashed curve) f0-curves of the four-syllable-long T2 sequences produced by the three speaker groups (with 95% confidence interval)
3.2. T4 sequences

In native speakers’ production, a clear pattern is observed in T4 sequences: the f0 curve of questions is realized in a higher frequency range, with elevated f0 compared to declaratives. According to the estimated means while T4Ds feature a decreasing tendency towards the end of the utterance, however T4 Qs are produced with a rising structure towards the end of the utterance (Table 3). Being more particular the average difference between estimated means of the moods are 2-3 semitones on the first two syllables, 4 semitones on the 3rd, and 8 semitones on the last syllable. Thus the difference between the moods gradually expands towards the end of the utterance (Table 3, Fig. 3). In the production of L2 learners, declining structure is observed in tendency for both Qs and Ds, and both groups carry out a raise on the ultimate syllable (compared to the penultimate). In the case of advanced learners’ production, the difference between the estimated means of the two moods are less than 1 semitone, except the ultimate syllable, where it reaches 2 semitones. In beginners’ production the distinction is usually less than 1 semitone, except for the second syllable. Regarding the smooth terms in native production, the curves’ shape differ between the two moods significantly in the case of all 4 syllable, furthermore, neither L2 group produces matching patterns with natives (except for beginners Ds in the 4th syllable). Regardless of intersections of Q and D curves in L2 production, we found that within each group the two modalities’ curves are discriminated significantly, Qs mostly yielding with higher values (except for when the two curves cross each other). However, in T4-production, in defiance of the falling characteristic of this tone, both Hungarian groups produce a strikingly different, rising curve, possibly due to preparing for the appropriate height for the fall of the subsequent syllable. As an example for tonal patterns, in case of the last two syllables of T4 sequences (especially the 3rd syllable, where the phase of the voiced initial is also included) a delayed target approximation is observed reaching the peak of the domed curve, as a transition realized connecting the preceding L target with the adjacent H [18]. In contrast, Hungarian production feature a moderately concave pattern in the 3rd syllable of T4 sequences. The difference between the two L2 groups’ strategy in the T4 sequence production is also apparent: while advanced learners realize both moods with similar patterns, beginners tend to be more inconsistent in the T4 production in Qs and Ds. It is also worth to mention that the T4’s pattern in the production of L2 learners always yield with a linear or concave pattern, as opposed to the native slightly domed curves. Regarding the effectiveness of the models signalled by the $R^2$ value, in each case, more than 90% of the dependent variables’ variance was explained by the analysed fixed and random effects.

Table 3: The parametric coefficients (the estimated mean (Est.), the t-value and the significance (above) and the smooth functions (the effective degrees of freedom (edf) and the F- and p-values) for the four T4 syllable, and $R^2$ for all models (below)

![Figure 3. The estimated question (red solid curve) and declarative (blue dashed curve) f0-curves of the four syllable-long T4 sequences produced by the three speaker groups (with 95% confidence interval)](image)

4. Conclusions

In sum, as expected, L2 learners did not approximate the elevated characteristic of natives’ question-production, however they did make distinction between the two moods by producing Qs slightly higher f0 than declaratives. We should further notice that our results confirm [4]’s results: there is a
significant range expansion on the last T2 syllable, while the T4 pattern remains roughly intact but elevated to a higher f0 range. While beginners approximated more native production by producing an apparent rising T2 with expanded range on the last syllable, in contrast, concerning the terminal T4 pattern, advanced learners approximated more the native production, not only elevating the Qs, but realizing the Qs and Ds with similar patterns as well. Comparing beginners’ and advanced learners’ production from a more global viewpoint, beginners were inconsistent in distinguishing interrogative and declarative patterns, while advanced learners tended to stick to the same f0 patterns in both, possibly because lexical tone-production schemes become practiced through the years of L2 experience, so in this manner advanced learners have more defined and intentional tonal realizations compared to beginners, whose production yields with less predictable patterns. However, regardless of interrogative or declarative mood, or the amount of experience, L2 learners’ tonal production yielded different patterns compared to natives, which is possibly due to fundamental problems in lexical tone production itself, as we can point out the concave T4 curves in contrast to the slightly domed native pattern, and the T2 patterns without rising phase. Our results could serve as a reference for a future experiment analysing spontaneous L2 dialogues. The results contribute to the understanding of the intonation acquisition of a tonal L2 language.

5. Acknowledgements

The authors are grateful to Tekla Etelka Grácz, Katalin Márda, Andrea Deme and Alexandra Markó for their kind suggestions and support. The research has been supported by UNKP-22-3 New National Excellence Program of the Ministry for Innovation and Technology, using resources of the Hungarian National Research, Development and Innovation Fund.

6. References