# The effect of explicit instruction and auditory/audio-visual training on Chinese learners' acquisition of English intonation

Yiling Chen, Ghada Khattab, Jalal Al-Tamimi

Newcastle University, UK
y.chen21@ncl.ac.uk, ghada.khattab@ncl.ac.uk, jalal.al-tamimi@ncl.ac.uk

## ABSTRACT

This paper examines the effectiveness of auditory vs. audio-visual training when coupled with explicit instructions on improving perception of intonation meanings by Mandarin-speaking learners of English as a foreign language (EFL). 40 subjects who were involved in the 3-week training process received instructions of intonation forms and functions based on the British approach. While half of them used Audacity (auditory-only) to manage the following self-paced practice, the other half used Praat (audio-visual). The perception experiment was administered by eliciting judgement of interpretations of contrastive intonation patterns with pre-, post-, and delayed post-test design, in comparison with two control groups (10 native RP speakers and 20 Mandarin speakers). It is found that auditory and audio-visual training had similar effect in facilitating perception of intonation meanings, indicating that visualisation does not necessarily improve the processing of intonation. This provides a reconsideration of the preference for audio-visual approach to intonation teaching.

**Keywords**: Intonation instruction, Audio-visual training, GLMM

## 1. INTRODUCTION

Intonation accounts for a big part of speech intelligibility and is notoriously difficult to be acquired by L2 learners. The bulk of research on L2 intonation has focussed on the examination of learners' intonational performance at the phonetic and phonological levels using perceptual and/or production tasks, e.g. [1], [2], [3]. However, little is known about how learners perceive semantic and pragmatic meanings of intonation. Without such knowledge, the underlying causes of the discovered difficulties at the phonetic and phonological levels are by no means clear.

The teaching of intonation has long been emphasised in research since the advent of Communicative Language Teaching (CLT) [4], but in practice, it is usually marginalised or even ignored [5]. The gap between research and practice might be attributed to a lack of empirical evidence from rigorously controlled experimentation of intonation training, and a failure of translations from intonation research to teaching materials appropriate for a particular group of EFL learners. Using speech tools to visualise pitch curves has been favoured when discussing intonation pedagogy as visual cues is argued to be facilitative to auditory cues [6]. However, research on the effectiveness of audio-visual training often failed to include an auditory-alone control group, e.g. [7], [8], leaving this argument less convincing. Aiming to fill these gaps, this study was dedicated to train on intonation with tailor-made instruction and practice materials specifically designed for Chinese EFL learners by addressing the following research questions.

First, can Chinese speakers distinguish interpretations of contrastive intonation patterns as native speakers do? Second, is intonation training effective in promoting learners' perception of intonation meanings? If yes, is there any difference of effects between auditory and audio-visual training?

## 2. METHODOLOGY

### 2.1. Participants

A total of 60 Chinese EFL learners aged between 19 and 34 participated in this project, 45 of whom (75%) were self-reported L1-Mandarin speakers, while the remaining claimed to have a good command of Mandarin though their L1 were local dialects. All of the participants were studying English-related postgraduate programs at Newcastle University, and had been learning English for more than six years.

Prior to the experiment, the Chinese participants were pseudo-randomly assigned to three groups, two experimental groups and one control group, according to their English proficiencies obtained from the Oxford Quick Placement Test (paper-and-pen version), and their latest IELTS scores. Statistical analyses of the between-group differences confirmed the homogeneity of Chinese groups in terms of the English proficiency at the starting point of the experiment.

Ten native self-claimed RP speakers were also recruited for the pre- and post-test to set a baseline for the analyses of perception tasks.

## 2.2. Experiment design

The experiment was done on week one (pre-test), week five (post-test), and week fourteen (delayed post-test). Following [9] and [10], twenty target utterances were selected from [11], [12] and [13] for the pre-test. Each of these utterances was generated with a pair of intonation patterns that convey distinctive meanings. Eight out of 20 utterances targeted at accentuation (tonicity), seven at prosodic phrasing (tonality), and five at nuclear tone. Another 20 utterances used in the post-test were all functionally matched to those in the pre-test to minimize a possible testing effect, while utterances used in the delayed post-test were the same as the post-test because little testing effect was expected for the two-month interval.

All the stimuli were recorded by two native speakers who had expertise in English phonetics via a Behringer ECM8000 microphone (15-20000 Hz frequency response) connected to the Edirol R-44 recorder (44.1 kHz sampling rate, 16 bits) in a soundproof booth of the speech lab at Newcastle University.

The experiment was administered in DMDX (5.1.3.4) in the Dell laptop. Each utterance was only repeated three times to control for the duration of the experiment and to minimize the learning effect, ended with 60 trials in total. In case the same utterance appearing after one another, the order of the trials was quasi-randomised and remained the same across participants. Within each trial, the test utterance followed by the two meaning glosses would appear on the screen for 10 seconds before playing the recording during when participants could familiarise themselves with the lexis. Once the recording was played, they were required to choose the correct interpretation of the meaning associated with the intonation pattern they heard by pressing the labelled key (A or B) corresponding to the meaning glosses. Three practice trials were included to ensure that participants were fully aware how to take the experiment.

## 2.3. Procedure of the experiment

The perception experiment was conducted in a soundproof booth in the speech lab. Participants came in individually at the time of their appointed slot. They were explicitly told that the experiment was about intonation. When anyone expressed uncertainty of what intonation was, very brief knowledge was verbally provided avoiding any technical term. A detailed explanation of the experiment had been provided before they settled themselves and adjusted the volume of the headphone (Bose QuietComfort 35) Once they pressed the space key to start, they were asked to follow the instruction on the laptop until they reached the end. Answers were automatically captured by DMDX and stored for later analyses.

## 2.4. Training procedure

The intonation training was done from week two to week five with six sessions each of which lasted about 2.5 hours with 60-75 mins of explicit instruction followed by 60-75 mins of self-paced practice. The instruction was based on [11] and [13], mainly focused on the informational, grammatical, and pragmatic functions of intonation realised by 3Ts (tonicity, tonality or tone) that are properly controlled at the phonetic level. Controversial usages and complex intonation patterns were omitted, so was the phonetic manipulations that Chinese EFL learners seem to have little difficulty with. The first session was centred on the general introduction to 3Ts and how important intonation was in communication. Session 2 focused on nuclear tones, session 3 and 4 on tonicity, session 5 on tonality, and the last session on all-together. The practice materials were based on the same sources, designed into activities focusing on particular intonation feature instructed in that session, in an order from shorter to longer utterances and to dialogues with a mixture of perception and production activities. Elicitation of linguistic knowledge of intonation always preceded the actual perception and production practice to make sure they understand why, which and how to use suitable intonation patterns to deliver intended meanings.

The only difference between the Audacity and the Praat group was during the practice, the latter could view the annotated pitch curves of the sample recordings and their own productions, whereas the former only accessed to auditory recordings.

# 3. RESULTS

## 3.1. Comparisons between native and Chinese participants for the pre-test

Before analysing the data relative to the research questions, we ruled out the items with the correct rate lower than 75% by native speakers and then checked the comparability of the pre- and post-test by running generalised linear mixed effect models (GLMM) [14] in R [15] (the effect of time was non-significant, $\chi^2(1)=2.81$, $p>0.05$).

In terms of the analyses of the pre-test, the Chinese participants were initially treated as one group to pinpoint the difference between them and the native speakers. The GLMM that fitted the data optimally included group, intonation feature, condition[1], the two-way interaction of condition and group, condition and feature, and group and feature as the fixed effects, with by-subject and by-item random intercepts as the random effects.

The results demonstrated that native listeners were in general significantly better than Chinese listeners in identifying intonation meanings ($b=-1.28$, $SE=0.34$, $z=-3.73$, $p<0.001$). By using the mixed () function in the afex package [16], a global effect of *group* was found on the overall judgement of the intonation meanings ($\chi^2(1)=87.10$, $p<0.0001$). Although *feature* had no main effect ($\chi^2(2)=0.40$, $p=0.82$), its interaction with *group* did ($\chi^2(2)=8.45$, $p=0.01$), suggesting that native and Chinese participants' discrimination of intonation meaning varied across different intonation features.

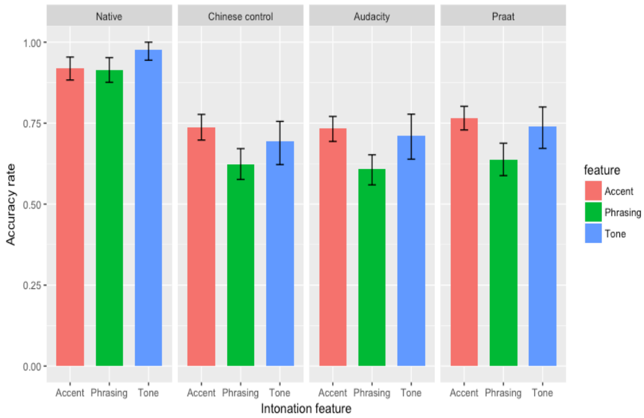
The post-hoc comparisons derived from the GLMM show that native speakers were significantly better than Chinese participants at all three intonation features (See Table 1). Among the Chinese group, the difficulty of understanding accentuation contrasts was no lesser than prosodic phrasing or nuclear tone contrasts.

**Table 1**: Comparisons between and within native and Chinese listeners on different features.

| Contrasts | Est. | SE | z | p | Sig. |
|---|---|---|---|---|---|
| Native accent--Chinese accent | -1.7 | 0.3 | -5.3 | <.001 | *** |
| Native phrasing--Chinese phrasing | -2.6 | 0.3 | -7.5 | <.001 | *** |
| Native tone--Chinese tone | -3.3 | 0.8 | -4.3 | <.001 | *** |
| Native accent--Native phrasing | 0.7 | 0.7 | 1.0 | 0.89 | N/A |
| Native accent--Native tone | 1.3 | 1.0 | 1.3 | 0.74 | N/A |
| Native phrasing--Native tone | 0.7 | 1.1 | 1.6 | 0.98 | N/A |
| Chinese accent--Chinese phrasing | -0.3 | 0.6 | -0.5 | 0.99 | N/A |
| Chinese accent--Chinese tone | -0.3 | 0.7 | -0.4 | 1.00 | N/A |
| Chinese phrasing--Chinese tone | -0.1 | 0.7 | -0.1 | 1.00 | N/A |

(Results were averaged over condition on the log odds ratio scale. P values were adjusted by mvt method.)

**Figure 1**: The mean accuracy of the comprehension task in the pre-test (averaged by condition).




### 3.2. Analysis of the training effect

For the purpose of investigating the training effect, the dataset of all three tests (pre-, post-, and delayed post-) by Chinese participants was targeted. The most fitted GLMM included six fixed effects and three random effects. The fixed effects were group (Chinese control, Audacity, and Praat), time, intonation feature, condition, the interaction of group and time, and the interaction of feature and condition. The random effects were the random intercepts for subjects and items, and the random slopes of feature for subjects.

The results show that *time* had a significant effect on the overall performance across groups ($\chi^2(2)=265.33$, $p<0.0001$), so did its interaction with *group* ($\chi^2(4)=161.89$, $p<0.0001$), indicating that the effect of time was different for each group. The significant effect was also found for group alone ($\chi^2(2)=54.45$, $p<0.0001$).

A series of subsequent GLMMs were run separately on the subset of the data for accentuation, prosodic phrasing and nuclear tone so that the post-hoc comparisons within each intonation feature can be deduced from these sub-GLMMs. Fig. 2 presents the predicted mean accuracy of accentuation across time, and it was statistically verified by the pairwise comparisons that the control group performed significantly worse than the two treatment groups in both the post- (control vs. Audacity, $ß=-0.9$, $p<0.001$; control vs. Praat, $ß=-1.01$, $p<0.001$) and delayed post-test (control vs. Audacity, $ß=-1.92$, $p<0.001$; control vs. Praat, $ß=-2.5$, $p<0.001$), while the difference between the two treatment groups was not significant.

[1] The interval variable "condition" had five levels: new1, new2, new3, old2, and old3. The number refers to the occurrence of the sentence, while "new" and "old" refers to the intonation pattern. So new1 refers to the first appearance of the sentence with a particular intonation pattern, and new2 refers to the second appearance of the same sentence but in the other intonation pattern. Old3 refers to the third repetition of the sentence spoken in an intonation pattern that was new, etc.

In addition, the Chinese control group did not improve at all from the pre- to the delayed post-test, while the two treatment groups did, as the difference between the pre- and delayed post-test was significant (for Audacity, ß=-1.7, $p<0.001$, for Praat, ß=-2.07, $p<0.001$) and so was the difference between the post- and delayed post-test (for Audacity, ß=-0.81, $p<0.01$, for Praat, ß=-1.28, $p<0.001$). Although the difference between the pre and post-test was marginal for the Audacity group (ß=-0.89, p=0.07) and non-significant (ß=-0.8, p=0.12) for the Praat group, it can be seen from Fig. 2 that the predicted accuracy rate increased from about 75% in the pre-test to nearly 90% in the post-test for both groups.

**Figure 2**: Predicted mean accuracy of accentuation for Chinese groups across time.

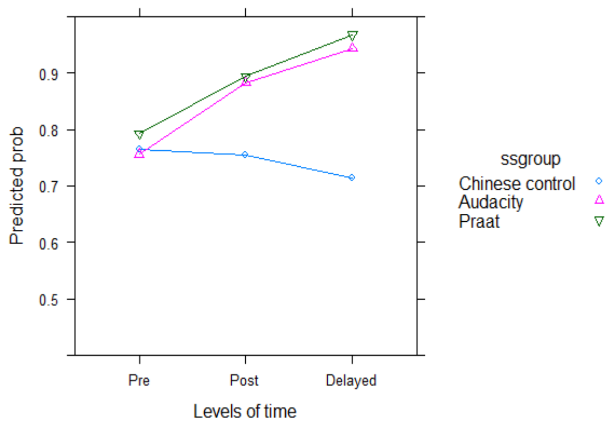



**Figure 3**: Predicted mean accuracy of prosodic phrasing for Chinese groups across time.

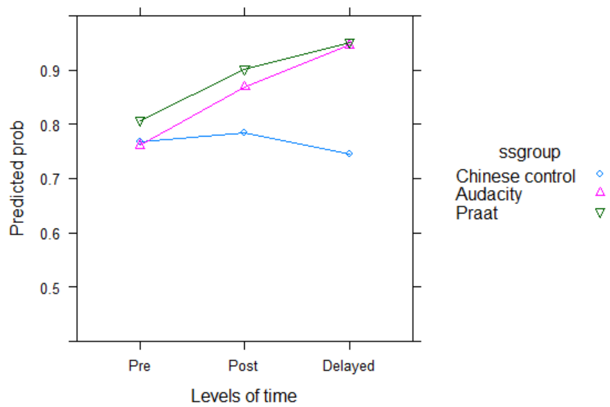
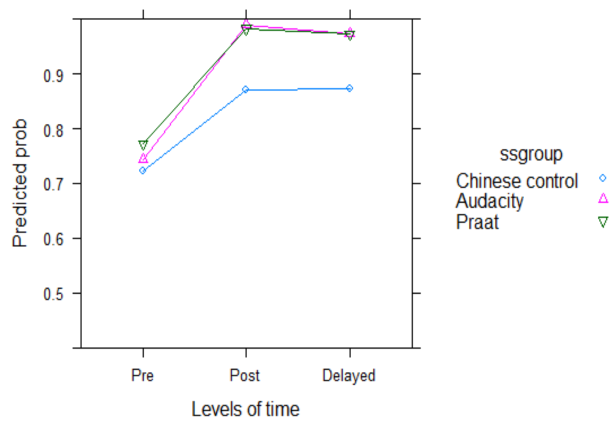



Fig. 3 shows the predicted mean accuracy of prosodic phrasing across time, and the pairwise differences between groups and between time points were the same as in accentuation in that the control group did not improve overtime, and it performed worse than the treatment groups at the post- and delayed post-test. For the treatment groups, no significant improvement was found between the pre- and post-test, but both were seen a significant improvement from the post- to the delayed post-test (for Audacity, ß=-0.97, $p<0.01$, for Praat, ß=-0.76, $p<0.05$).

Results from the pairwise comparisons for nuclear tone (Fig. 4) were slightly different from those of prosodic phrasing and accentuation, as from the pre- to the post-test both treatment groups showed a significant improvement (for Audacity, ß=-3.44, $p<0.001$, for Praat, ß=-2.84, $p<0.001$). The difference between the post- and delayed post-test was not significant due to the ceiling effect. For the control group, the observed improvement from the pre- to the post-test turned out to be non-significant (ß=-0.09, p=0.99).

**Figure 4**: Predicted mean accuracy of nuclear tone for Chinese groups across time.




## 4. CONCLUSION

This study has found that Chinese EFL learners were significantly worse than native speakers in identifying semantic and pragmatic meanings of English intonation encoded by accentuation, prosodic phrasing, and nuclear tone, and that their difficulties in understanding intonation features were equally scaled. In regard to the training effect, learners' comprehension ability was improved immediately after the training for all three features, and continued to improve as evidenced in their native-like performance in the delayed post-test, indicating that certain aspects of intonation are teachable and learnable, and tailor-made instruction and materials are effective and applicable in use. More interestingly, the auditory group performed and improved in a similar fashion as the audio-visual group, suggesting that audio-visual training does not outperform auditory-alone training for the teaching and learning of intonation meanings.

## 5. REFERENCES


[1] Ding, H., O. Jokisch, and R. Hoffmann, *A Phonetic Investigation of Intonational Foreign Accent in Mandarin Chinese Learners of German*, in *Speech Prosody* 2012.

[2] Li, A. and B. Post, *L2 acquisition of prosodic properties of speech rhythm.* Studies in Second Language Acquisition, 2014. **36**(02): p. 223-255.

[3] Liu, X. and X. Chen, *The acquisition of English pitch accents by Mandarin Chinese speakers as affected by boundary tones*, in *Speech Prosody*. 2016: Boston, USA. p. 956-960.

[4] Derwing, T.M. and M.J. Munro, *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Vol. 42. 2015, Amsterdam, Netherland: John Benjamins Publishing Company.

[5]Wang, G., *Guanyu yingyu yuyin kecheng jianshe de jidian sikao [A few constructive comments on the development of English pronunciation curriculum]*, in *Zhonguo Eryu Yuyin Xide Yanjiu De Xianzhuang Ji Fazhan Qushi [Second Language Phonological Studies in China]*, H. Chen and J. Yang, Editors. 2010, Foreign Language Teaching and Research Press: Beijing, China. p. 15-31.

[6]Hardison, D.M., *Generalization of computer-assisted prosody training: Quantitative and qualitative findings.* Language Learning & Technology, 2004. **8**(1): p. 34-52.

[7] Gorjian, B., A. Hayati, and P. Pourkhoni, *Using Praat software in teaching prosodic features to EFL learners.* Procedia - Social and Behavioral Sciences, 2013. **84**(0): p. 34-40.

[8] Le, H.T. and J. Brook, *Using Praat to teach intonation to ESL students.* Hawaii Pacific University TESOL Working Paper Series, 2011. **9**(1, 2): p. 2-15.

[9] Cruz-Ferreira, M., *Non-native interpretive strategies for intonational meaning: An experimental study*, in *Sound Patterns in Second Language Acquisition*, A. James and J. Leather, Editors. 1987, Foris: Dordrecht. p. 103-120.

[10] Atoye, R., *Non-Native perception and interpretation of English intonation.* Nordic Journal of African Studies, 2005. **14**(1): p. 26-42.

[11] Wells, J.C., *English Intonation: An Introduction*. 2006, Cambridge: Cambridge University Press.

[12] Bu, Y., *English Intonation: Its Form, Function and Application*. 2003, Beijing: Foreign Language Teaching&Research Press.

[13] Carr, P., *English Phonetics and Phonology: An Introduction*. 2nd ed. 2013: Wiley-Blackwell.

[14] Bates, D., et al., *Fitting Linear Mixed-Effects Models Using lme4.* Journal of Statistical Software, 2015. **67**(1): p. 1-48.

[15] Team, R.C., *R: A language and environment for statistical computing*. 2014, R Foundation for Statistical Computing, Vienna, Austria: Vienna, Austria.

[16] Singmann, H., et al. *afex: Analysis of Factorial Experiments. R package version 0.16-1*. 2016; Available from: https://CRAN.R-project.org/package=afex.