Using Multimodal Methods in L2 Intonation Teaching for Chinese EFL Learners

Chenyang Zhao
University of Chinese Academy of Social Sciences
Beijing, China
cyzhao16@mails.jlu.edu.cn

Ziyu Xiong
Institute of Linguistics
Chinese Academy of Social Sciences
Beijing, China
xiongziyu@163.com

Aijun Li
Institute of Linguistics
Chinese Academy of Social Sciences
Beijing, China
liaj@cass.org.cn

***Abstract*——The present study aims to investigate how multimodal training method contribute to the improvement of the L2 intonation produced by Chinese EFL learners. Altogether 75 learners with an English major background from 3 different dialectal regions of China are recruited. They are divided into 5 groups which differ from each other in training methods, which specifically are the control group (G1), group with sound for training only (G2), group with sound and after-training feedback (G3), group with both audio and visual material for training (G4), and the audiovisual training group with feedback (G5). The results show that although no significant improvement between learners' pretest and posttest for each group, still we observe that some of the learners in experiment groups score significantly higher in posttest than those in the control group, and among them, G5 is the best as the most cases of intonation are improved through the training. This indicates that multimodal + supervised training method is the most effective way in L2 intonation teaching in this experiment. Unobvious improvement of in the rest cases might due to the limited training time, which will be further ameliorated by a supplementary intensive training in this method.**

Keywords—multimodal training method, feedback, English L2 intonation, Chinese EFL learners

I. INTRODUCTION

A. Studies on L2 English intonation of Chinese learners

Intonation of an utterance has long been recognized as an important cue helping to get the intention of utterers understood. Deviance in intonation might not only lead to foreign accent, but also impair the intelligibility and compressibility of sentences, particularly in L2 communication [7]. For these reasons, relevant pedegogic research has been conducted over the issue. However, it seems that English learners from China do not demonstrate a satisfying performance in this aspect. According to the experimental research of Ji [14], Chinese learners tend to apply falling tone in yes-no questions, which is different from the native speakers' rising tone in realizing the nuclear accent. Wang [15] also finds that for yes-no questions of English, Chinese learners show an instability of selection in pitch pattern of nuclear accent and boundary tone application, specifically by switching among L*H, H*, or H*L and L% or H% respectively, without a regular pattern to follow. Apart from this particular sentence type, deviations can also be found from the L2 intonations of other types. Li et al [1] discovered that Chinese learners from Jinan dialectal area do not use the same prosodic cues to relization the narrow focus in English statements as the native speakers do. In addition, learners from the same dialectal area are not able to convey different semantic meanings of tag questions through the employment of the correponding tones, instead, they take this sentence type as a simple combination of statement and yes-no question, and apply the prototypical tones of the two to do the productions, without any variation according to the intentions of the particular utterance [12]. As to complex sentence types such as complentary sentences, Chinese learners from Northeastern dialectal area are good at falling tone relization, but still find it difficult to produce sentences with level tone [16].

B. Application of Multimodal methods in L2 education

All of these deviated cases in L2 production indicates that a new way is supposed to be found in the education of Chiniese learners' L2 intonation. Multimodal teaching in recent years is an emerging method applied in teaching process by some practitioners. With both audio and visual information presented simultaneously, learners are expected to produce a less deviant sentence from the native speakers. By investigating features of L2 Japanese produced by Chinese learners, Liu [9] produced that with the help of spectrogram in praat, figures of articulators and lips, and the MRI-based amination of articulation in teaching, an improvement is made in learners' productions of devoiced vowels, voice and voiceless stops, apical flaps and nasals, and syllable-final nasals. By the application of this method, learners are observed to have developed an ability of self-monitoring and self-feedback in their pronunciation. A 3D feedback system with a higher automatibility is designed by Katz et al. [13] to show the real-time tongue and jaw movement in the production of each segment. Everytime learners' articulators are in the correct place, the system would provide a positive feedback, by which learners are able to acquire the right way of articulation after long time of training. In suprasegmental level, the feasibility of multimodal teaching is explored as well. Zhuang and Bu [10] discover that, BetterAccent Tutor, a visual intonation learning software, can help learners to better locate the lexcial stress in word level, or to help distinguish the stressed-unstress syllables in words. This software is designed to have an utterance visualized so that learners can visually perceive the length, pitch, and intensity of vowels [10]. But such a software still have not been generalized in sentence-level intonation teching. Demenko et al [5] have the gap filled by their audiovisual prosody training system which, specifically, present both the native and the learners' pitch contours simultaneously in a visual way, so that the latter's attention could be attracted to voluntarily compare their falling or rising slopes with the formers. This system is proved effective among German learners with an L1 Russian background and vice versa. Lee et al. [5] make the feedback automatized through the CALL system so that learners who are not sensitive to their deviance from the native speakers could be provided an automatic warning. The system is tested among English learners from Korea. With the function of stress prediction and detection, learners can clearly notice whether they put the stress in the correct place, and if they realize it in the proper pitch pattern without any delay. The results of this experiment show that learners' accuracy on ryhthm and fluency are improved after training by this system [6].

All the studies cited here suggest that multimodal methods are conducive to L2 intonation training. However we observe that none of the learners participated in experiments listed above are informed with the relevant linguistic knowledge on intonation, therefore it is doubted that if the effect acquired in real time would last long, and without the aid of the system, whether or not they would fall back to their original way of production in real-life communication. Given that we decide to use a semi-automatic audiovisual system in present study to see how it works in English intonation training among learners from China.

II. METHODS

A. Experiemtal sentence

In present study, we determine to focus on the 3 typical sentence types in English, statement, yes-no question, and WH-question. Altogether 5 different teaching methods are used to see if learners could have an improvement in intonation of the above 3 sentence types. Presented in the next table are the experimental sentences used in the pre- and post-test of the research.

TABLE I. EXAMPLES OF EXPERIMENTAL SENTENCES

Sentence type	*Focus Placement*	*Inducing sentence*	*Target sentence*
statement	initial	Who plays the clarinet in the class at school?	DON plays the clarinet in the class at school.
statement	medial	What does Don play in the class at school?	Don plays the CLARINET in the class at school.
statement	final	Where does Don play the clarinet?	Don plays the clarinet in the class at SCHOOL.
statement	initial	Who left the socks on the fence?	MARK left the socks on the fence.
statement	medial	What did Mark leave on the fence?	Mark left the SOCKS on the fence.
statement	final	Where did Mark leave the socks?	Mark left the socks on the FENCE.
Yes-no	initial	Anna comes to the party with Tom.	Can JANE come with Tom?
Yes-no	medial	Guess where did the ship depart in the afternoon?	Did the ship depart from JAPAN in the afternoon?
Yes-no	final	Jane comes to the party with Manny.	Can Jane come with TOM?
Yes-no	initial	Saran comes to the party with Burnell.	Can CATHERINE come with Burnell?
Yes-no	medial	Guess from where did the ship departed in the morning?	Did the ship depart from JAPAN in the morning?
Yes-no	final	Guess when did the ship depart from Germany?	Did the ship depart from Germany in the AFTERNOON?
WH	initial	The ship departed from France on Sunday.	WHAT departed from France on Sunday?
WH	medial	It's the absolute truth, I swear it.	But who's going to BELIEVE such a fantastic story?
WH	final	What was that you said?	Where did you go for your summer HOLIDAY?
WH	initial	The ship departed from Japan in the afternoon.	WHEN did ship departed from Japan?
WH	medial	I haven't seen Jenny for ages.	Who was that you were talking to in the STREET this morning then?
WH	final	I'm very fond of Jellied eel.	How in the world can you eat such STUFF?

Altogether 18 sentences are cited each time for doing this teaching experiment, part of which are cited from O'Connor and Frederick's [8] *Intonation of Colloquial English*, and the rest used to serve in the establishment of the database AESOP-CASS [17]. Each one is composed of an inducing sentence and a target one, the capitalized word in which is the semantic focus.

B. Equipment

Due to the COVID-19 in this year, the whole experiment is conducted on line through Tencent Meeting, on which a small-scale on-line classroom teaching is realized. X-perception [18], a PC-end software is employed in the process to do material transmission between learners and experimenters. This is a platform able to simultaneously demonstrate text, audio, and visual materials. Experimenters are allowed to adjust the play times of each audio track and the intervals between two plays. Besides, it also works as a real-time recorder, by which learners can have their productions recorded and automatically stored in a specific folder for experimenters' reference. Presented below is the interface of X-perception.

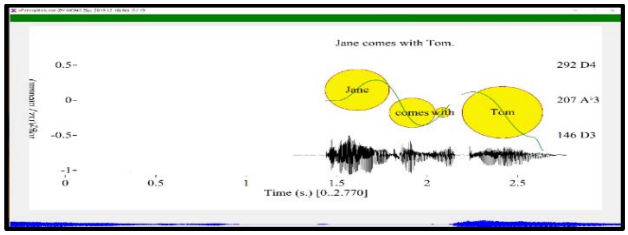

Fig. 1. Interface of X-perception

To draw the visual intonation contour as is displayed in Fig. 1, ProZed, a Praat-based tool designed by Hirst [2] [5] is employed, by which the duration of each syllable, pitch movement, and intensity of an utterance are visualized. The visual pitch contour is drawn by using the OMe (Octave-Median) scale [3]. The Octave-Median scale uses the median pitch as a central value and displays the pitch in octaves with respect to this pitch. Typical unemphatic pitch tends to be centered within one octave centered around the median value. Syllables within the utterance are displayed as circles, the diameter of the circle corresponds to the duration of the unit, and the height of the circle corresponds to the mean pitch of the syllable. Specifically, the green curve in Fig.1 represents the movement of pitch contour, and the yellow circle the syllables. Inside each circle the corresponding text of the syllable is written, which is the same as that in the sentence displayed at the top.

C. Participants and experiment process

a) Participants

To test the feasibility of this audiovisual system in intonation teaching, we recruit altogether 75 participants whose group distribution is presented below.

TABLE II. DISTRIBUTIONS OF EXPERIEMT PARTICIPANTS

Training method / Dialectal area	*G1*	*G2*	*G3*	*G4*	*G5*
Jianghuai		4	5	5	5
Beijing	5	5	5	5	5
Northern	4	5	5	5	5

It is noticed from the table that participants are from 3 different dialectal areas of China, Jianghuai dialectal area, Beijing dialectal area, and Northern dialectal area. All of them are undergraduate English majors aged from 18-22 years old, born and raised in their hometown, and able to communicate in both their mother tongue dialect and Mandarin Chinese. None of them are reported to have hearing or sight impairment. Participants in each dialectal group are then subdivided into different training groups, where they are given the same learning materials but are trained through 5 different methods.

- Group 1: This is the control group and no training is provided for them who however is not prohibited to learn intonation through their own way during the experiment;
- Group 2: Learners in this group are trained only by audio materials. They are asked to listen the native

speakers' productions and try to imitate their intonation in practice;

- Group 3: Like the above group, what they use for training is also the audio tracks of native speakers' utterance, but each time after training they would be given a feedback by experimenters according to mistakes they commit in practice;
- Group 4: Visual intonation contour in this group is used, together with audio productions of native speakers. Learners can listen and observe the contours at the same time to discover features of native speakers' intonation;
- Group 5: Also like group 4, they are provided with both audio and visual materials for imitation. Additionally, after-training feedback are also provided to help them better understand their deviance from the native speakers.

It is noted that learners in group 3 and group 5 are equipped with the basic linguistic knowledge on English intonation. Experimenters would organize an on-line class on Tencent Meeting right ahead of each training. They present features of the 3 different sentence types in the aspects of nuclear pitch pattern, boundary tone, and variation of duration. The predicated deviations are also explained to learners according to the previous study on L2 intonation. This is done in order to help them develop a "keen eye" in training to better perceive the minute variation of intonation. But this pre-training teaching is different between the above two groups in methods, with group 3 using audio materials only and group 5 the visual intonation contour as well, and the distinction in teaching method is still preserved in the process of after-training feedback.

b) Experiment process

Followed next are the exact steps of this experiment. The whole process have lasted for 16 days with 5 times of training, a pretest and a posttest. Trainings are conducted each time with a one-day interval.

- Day 1: Pretest for all the participants;
- Day 4: Linguistic knowledge explanation for group 3 and 5; first training for group 2-4.This is done after 2 days of pre-test;
- Day 6: Feedback for group 3 and 5, and new linguistic knowledge preparation for them; second training for group 2-5;
- ……
- Day 13: Fifth training for group 2-5, with the post-training feedback for the fourth training of group 3 and 5;
- Day 16: Posttest for all the participant 2 days after the final training.

OFF-LINE TRAINING

In each training day, learners would receive a folder with 18 sentences which are displayed separately in different pages. Learners can hear both the inducing and target sentences, but are only provided the target one in visual and textual way, so that their can focus their attention on grasping the target sentence' intonation. All the sentences are read by native English speakers from Middle or West America. Learners need to listen each of them for 3 times, and try to observe the features from the audio and visual intonation contours (if they have in their folders). After that they are required to repeat and record the same sentences. This software allows endless recording of the same utterance in case learners are not satisfied with productions. The training is conducted off line and in a self-monitoring way. Each learner is asked to complete the imitating and recording tasks within 40-50 minutes in same the training day, after which the folders with their own recorded productions are supposed to be submit to experimenters for rating and feedback preparation (group 3 and 5 only).

Sentences used for training in each time are different. Table 2 displays their distribution in focus and sentence type.

TABLE III. DISTRIBUTION OF EXPERIMENTAL SENTENCES IN TRAININGS

	statement			**Yes-no**			**WH**		
	initial	***medial***	***final***	***initial***	***medial***	***final***	***initial***	***medial***	***final***
T 1	6	6	6						
T 2				6	6	6			
T 3							6	6	6
T 4	2	2	2	2	2	2	2	2	2
T 5	2	2	2	2	2	2	2	2	2

It is noticed from the table that the first 3 times of trainings are focused on statement, yes-no question, and WH-questions respectively, while in the last two times, the 3 sentence types are mixed together for them to practice, by which learners are expected to have a review of what they learned previously.

RATING

Having collected the productions of learners in each training, of which, group 3 and group 5's would be rated by experimenters and on this basis, the feedback of their performance is given through the comparison of their production and the native speakers'. Specifically, falling (H*L), rising-falling (L*HL), high level tone (H*) are equally used by native speakers statements, and as to yes-no question, L*H appears frequently in those with sentence-medial focus, and L*H in the sentence-final ones, and for wh-question, H*L and H*L (or L*HL) are applied respectively; in addition, they realize the boundary tones of statement and wh-question by H% and yes-no L% in most cases [14] [15]. Learners whose productions do not fall in one of the cases would be defined false, and otherwise correct.

TABLE IV. EXAMPLE OF RATING CHART

Subject 1	***Nuclear accent placement***	***Pitch pattern of focus***	***Boundary tone***
S1	1	0	1
S2	0	1	1
S3	0	0	0
S4	1	1	1
S5	1	0	0

S6	0	1	1

Presented above is the rating chart used by experimenters in each time. We notice that learners are rated from 3 perspectives, nuclear accent placement, pitch pattern of focus, and boundary tone. "1" represents a correct realization, and "0" the false. The results are used to do feedback for group 3 and 5.

Rating method employed is referred from the criteria of Minematsu et al [11] with some adaptions made to meet the requirement of this experiment. Segmental, rhythmic, and intonational realizations of a sentence are respectively rated by a 5-scale criteria as a whole, but here in present study, only the learners' representation of intonation is focused, and its performance is further divided into 3 different aspects which can be judged correct or not in a clear-cut manner without considering the exact degree. Therefore, the 5-scale rating criteria is abandoned.

FEEDBACK

For learners in group 3, the incorrect productions are compared with the native speakers' audio tracks. These are demonstrated one by one in slides of PowerPoint in order to help them locate the mistakes and how they are defined wrong to facilitate their self-correction and future training. For Group 5, the feedback is conducted with the aid of both sound and visual intonation.

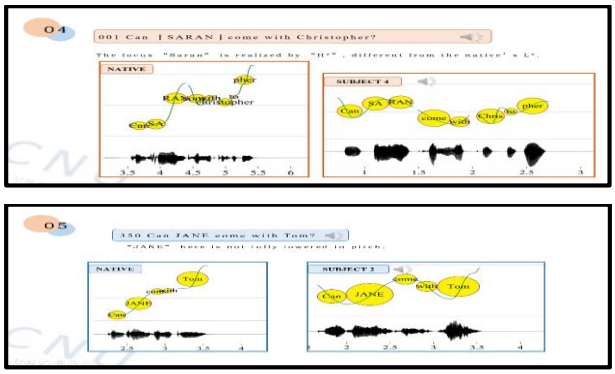

Fig. 2. Examples of Feedback for Group 5

The entire on-line feedback process last for 30-40 minutes each time. Through the observation of the figure and sound, together with the experimenters' explanations, learners are supposed to have the same types of mistakes avoided in the posttest and their own reading practice.

III. RESULTS

Upon completion of whole process, experimenters start to have their productions in pretest and posttest rated in accordance with the same standard used in training. The results are discussed in this section. As 7 learners stopped training in middle of experiment out of their personal factors, only 68 of them are rated.

As is mentioned above, 6 sentences are assigned to each sentence type in the pretest and posttest. Each learner would score from 0-6 in an individual aspect and 0-18 in total. The average scores of all the learners in the same group taken as the representative to do the following statistical analysis.

We first compare overall scores of learners in each group to see how they progress between the pretest and the posttest.

TABLE V. MEAN SCORES OF 3 SENTENCE TYPES IN PRE-POST TESTS

Training method / Dialectal area	G1		G2		G3		G4		G5	
test	pre	post	pre	post	pre	post	pre	post	pre	post
Jianghuai			25.4	30	29.6	46.5	33	35.4	34	35.8
Beijing	36.6	37.4	31.6	37.2	35.4	39.8	22.8	34.8	27.6	38.2
Northern	32.8	36	30.2	38.2	35.4	40.4	32.6	34.6	38.2	43.8

TABLE VI. PARI-WISE T-TEST OF 3 SENTENCE TYPES IN PRE-POST TESTS (P VALUE)

Training method / Dialectal area	*G1*	*G2*	*G3*	*G4*	*G5*
Jianghuai		.139	.135	.592	.177
Beijing	.721	.103	.317	**.023**	**.002**
Northern	.51	.027	.119	.75	.175

From the above two tables, we notice that learners in all the groups have made a better performance in their posttest than the pretest, however, the progress they have obtained during the training is not significant enough except two groups marked in bold in Table 6. We doubt that the training methods employed in this experiment is only conducive to the particular aspects of intonation teaching.

TABLE VII. MEAN SCORE OF 3 SENTENCE TYPES IN PRE-POST TESTS (P VALUE)

Northern Dialect	Statement						Yes-no						WH					
	Nuclear accent Placement		*Pitch pattern*		*Boundary tone*		*Nuclear accent Placement*		*Pitch pattern*		*Boundary tone*		*Nuclear accent Placement*		*Pitch pattern*		*Boundary tone*	
test	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post
G1	3.5	1.8	4.3	4.5	6	6	2.3	2.5	1.5	5	5.5	5.8	1.3	1.8	3	4	5.5	4.8
G2	2.6	4	4.2	5	5.8	6	2.2	3.6	2.4	5.6	4.6	5.4	1.6	1.8	3.6	3.4	3.2	3.4
G3	4.2	4.8	4.8	5.6	5.8	6	3.6	4.6	2.8	3	4.4	5.6	1	1.6	3.4	4	5.4	5.2
G4	3.2	2.8	5.2	5.2	6	6	2	1.6	1.6	3	3	4.6	1.4	0.8	5	5	5.2	5.6
G5	4.4	6	5.6	6	6	5.8	5.8	5.4	1	4.6	5.2	5.8	2	3	3.2	2.8	5	4.4

TABLE VIII. PARI-WISE T-TEST OF SCORES OF PRE-POST TESTS IN 3 ASPECTS

Northern Dialect	Statement			Yes-no			WH		
	Nuclear accent Placement	*Pitch pattern*	*Boundary tone*	*Nuclear accent Placement*	*Pitch pattern*	*Boundary tone*	*Nuclear accent Placement*	*Pitch pattern*	*Boundary tone*
G 1	.102	.718	1.000	.391	**.006**	.391	.391	.092	.215
G 2	.080	.242	.374	.025	**.003**	.099	.621	.621	.871
G 3	.305	.099	.374	.351	.749	.109	.208	.374	.621
G 4	.587	1.000	1.000	.178	.108	.140	.374	1.000	.374
G 5	.078	374	.374	.477	**.018**	.305	.142	.688	.634

We take results of learners from Northern dialectal area as an example to see how the training methods contribute to different aspects of intonation amelioration in the 3 sentence types. The comparisons made in both Table 7 and Table 8 show that although most of the groups score higher in their

posttest of all the 3 intonation aspects, still the enhancement is not significant enough except in 3 cases marked in bold in Table 7. However, according to previous studies, multimodal method is proved to be effective in intonation teaching. It might be due to the limited time of training in this experiment that leads to the unobvious improvement. But as equal time is assigned to each group, it is worthwhile to see if there is a significant difference of posttest scores among them in 3 aspects of intonation in all the sentence types.

A. Nuclear accent Placement

In this section, scores of posttest of the 4 multimodal training groups are compared with those of control group to see which of them are more effective in the improvement of intonation. We start from the aspect of nuclear accent placement. As all participants in group 1 of Jianghuai Dialect quit the experiment, only learners from Northern and Beijing Dialectal Areas are compared in this way.

TABLE IX. INDPENDENT T-TEST OF SCORES BETWEEN CONTROL AND EXPERIMENT GROUPS IN NUCLEAR ACCENT PLACEMENT POSTTEST (P VALUE)

Northern Dialect	***G1-G2***	***G1-G3***	***G1-G4***	***G1-G5***
Statement	**.001**	**.000**	.175	**.000**
Yes-No	**.022**	.050	.255	**.001**
WH	.879	.801	**.020**	**.021**
Beijing Dialect	***G1-G2***	***G1-G3***	***G1-G4***	***G1-G5***
Statement	.347	.771	.475	1.000
Yes-No	1.000	.207	**.034**	**.008**
WH	.095	.694	**.002**	**.243**

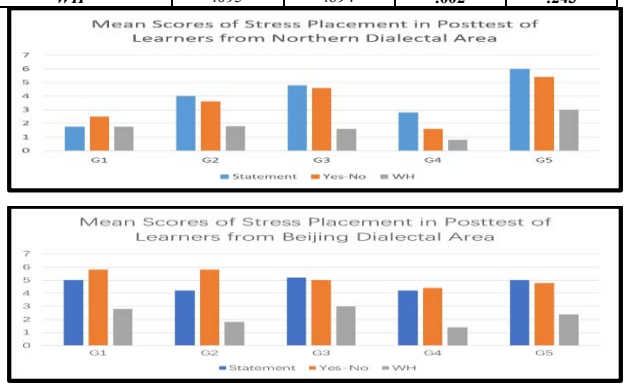

Fig. 3. Means scores of nuclear accent placement in posttest of learners from Northern and Beijing dialectal areas

According to the comparison, it is noticed that learners in group 5 of Northern dialectal area score significantly higher than their counterparts in group 1 for all the sentence types; besides, group 5 and 3, the only two groups with post-training feedback, achieve the first and the second highest scores in this aspect among all the learners.

However, result of learners from Beijing dialectal area goes the opposite to their Northern fellows. group 5 is significantly lower than group 1 in the nuclear accent placement for all the sentence types, so is group 4, another one with visual intonation contour available in training. group 2 and 3, the two groups with only audio productions of native speakers, though obtain higher scores than the control group in some cases, still is not significant enough.

B. Pitch Pattern of Nuclear Accent

TABLE X. INDPENDENT T-TEST OF SCORES BETWEEN CONTROL AND EXPERIMENT GROUPS IN PITCH PATTERN OF POSTTEST (P VALUE)

Northern Dialect	***G1-G2***	***G1-G3***	***G1-G4***	***G1-G5***
Statement	.571	.125	.356	.103
Yes-No	.407	**.027**	.073	.667
WH	.334	1.000	**.027**	**.033**
Beijing Dialect	***G1-G2***	***G1-G3***	***G1-G4***	***G1-G5***
Statement	.143	.694	.318	.771
Yes-No	**.011**	.093	.227	**.044**
WH	.305	.545	**.020**	1.000

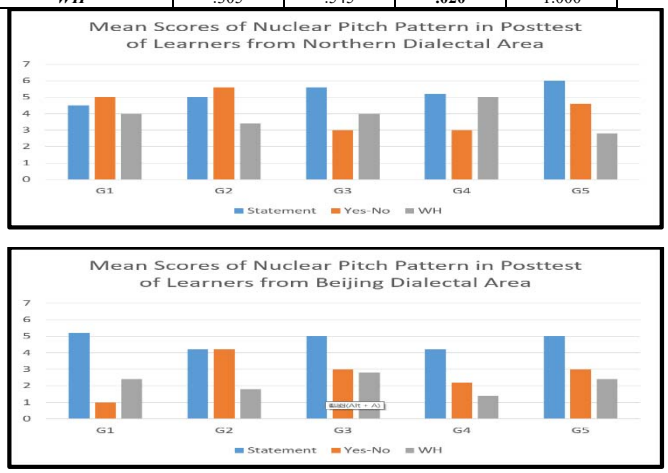

Fig. 4. Means scores of nuclear pitch pattern in posttest of learners from Northern and Beijing dialectal areas

The comparison made above is on the representation of focus pitch pattern, by which it is observed that all the training methods do not contribute to the improvement of statement production in this aspect as no significant difference is found between scores of these 4 groups and the control group. But for the other two sentence types, each training method works differently, with the audiovisual methods used in group 4 and 5 achieving the best performance by enhancing the wh-question production of learners from the two dialect areas, and the yes-no question productions of learners from Northern dialectal area and wh-question productions of those from Beijing.

C. Boundary Tone

An excellent style manual for science writers is [7].

TABLE XI. INDPENDENT T-TEST OF SCORES BETWEEN CONTROL AND EXPERIMENT GROUPS IN BOUNDARY TONE OF POSTTEST (P VALUE)

Northern Dialect	***G1-G2***	***G1-G3***	***G1-G4***	***G1-G5***
Statement	1.000	1.000	1.000	.407
Yes-No	.639	.775	.419	.879
WH	.224	.539	.273	.777
Beijing Dialect	***G1-G2***	***G1-G3***	***G1-G4***	***G1-G5***
Statement	.178	.178	.141	.545
Yes-No	.208	.733	.667	.397
WH	.239	.870	.226	.665

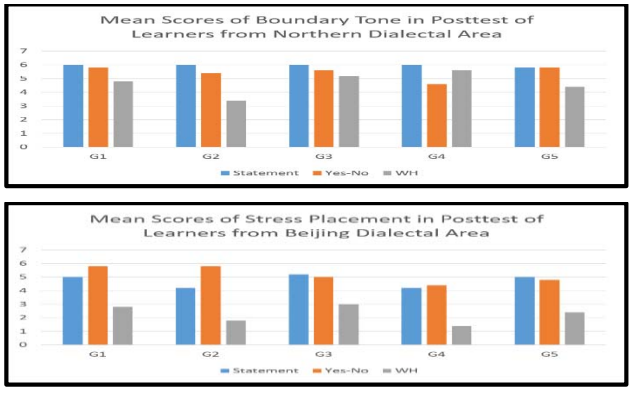

Fig. 5. Means scores of boundary tone in posttest of learners from Northern and Beijing dialectal areas

Obviously in Table 11 and Fig. 5, no significant difference is found between learners from experiment groups and control group in the realization of boundary tone. Learners in all the groups score relatively higher (approximately 6) in this aspect of statements and yes-no questions, so are wh-questions produced by those from Northern dialectal area. Learners with L1 Beijing dialectal in different groups obtain a relatively equal score of around 3.

IV. Discussions and Conclusions

Effect of the 4 training methods in all the sentence types of 3 different intonation aspects are concluded as follow. Marks of " ↑ ", " ↓ ", and "-" respectively represents "significantly higher", "significantly lower", and "no significant difference" compared with scores of the control group.

TABLE XII. Indpendent T-Test of Scores between Control and Experiment Groups in Boundary Tone of Posttest (P Value)

Nuclear accent placement	***Northern Dialect***	***G2***	***G3***	***G4***	***G5***
	Statement	↑	↑	-	↑
	Yes-No	↑	-	-	↑
	WH	-	-	↓	↑
	Beijing Dialect	***G2***	***G3***	***G4***	***G5***
	Statement	-	-	-	-
	Yes-No	-	-	↓	↓
	WH	-	-	↓	↓
Pitch pattern of focus	***Northern Dialect***	***G2***	***G3***	***G4***	***G5***
	Statement	-	-	-	-
	Yes-No	-	↓	-	-
	WH	-	-	↑	↓
	Beijing Dialect	***G2***	***G3***	***G4***	***G5***
	Statement	-	-	-	-
	Yes-No	↑	-	-	↑
	WH	-	-	↓	-
Boundary tone	***Northern Dialect***	***G2***	***G3***	***G4***	***G5***
	Statement	-	-	-	-
	Yes-No	-	-	-	-
	WH	-	-	-	-
	Beijing Dialect	***G2***	***G3***	***G4***	***G5***
	Statement	-	-	-	-
	Yes-No	-	-	-	-
	WH	-	-	-	-

Boundary tones of all the sentence types are not significantly improved among all the experiment groups in 3 sentence types, which might be due to the relatively good performance of this aspecct in pretest, leaving less room for learners to improve. Considering the other two aspects, training method used by group 5 seems to be the most effective, as altogether 4 cases are observed higher than the control group, especially in nuclear accent placement. Cases marked by " ↓ " though is significantly lower than the control group, still is not sigficantly different from the pretest performance of their own. The unonvious progress might be caused by the limited training time. But this also indicates that the trainings conducted in this experiment at least impose little negative effect on learners' L2 intonation.

Through the above discussion, we can draw a preliminary conclusion that audiovisual training (native sound+viusal intonation contour) with a pre-class linguistic knowledge preparation and an after-class feedback is the most effective way to enhance Chinese learners' production of L2 English intonation. A supplementary intensive training is supposed to be done later by this method to gain a more obvious and stable progress in L2 intonation.

Acknowledgment

This work is supported by National Major Social Sciences Foundation of China under Grant 15ZDB103, and Innovation Program of Chinese Academy of Social Sciences. Special thanks to Xiaoli Ji, Na Zhi, and Liang Zhang for providing valuable suggestions to this experiment.

References

[1] A. Li, X, Wan, C. Zhao, & L, Zhu. "Phonological and phonetic realization of narrow focus in declarative sentences by Jinan EFL learners", In Proceedings of Speech Prosody 2020.

[2] B. Bigi, & D. J. Hirst. "Speech phonetization alignment and syllabification (SPPAS): a tool for the automatic analysis of speech prosody". In Proceedings of the 6th International Conference on Speech Prosody, 2012.

[3] C. D. Looze, & D. J. Hirst. "The OMe (Octave-Median) scale: a natural scale for speech prosody", In N. Campbell, D. Gibbon, & D.J. Hirst, eds, Proceedings of the 7th International Conference on Speech Prosody, Trinity College, Dublin, Ireland, May 2014.

[4] D. J. Hirst. "ProZed: a speech prosody editor for linguists, using analysis-by-synthesis", In Keikichi Hirose; Jianhua Tao (eds.) Speech Prosody in Speech Synthesis. Modeling and Generation of Prosody for High Quality and Flexible Speech Synthesis. (in series Prosody, Phonology and Phonetics), Berlin, Heidelberg: Springer Verlag. 2015, pp. 3-17.

[5] G. Demenko, A. Wagner, N. Cylwik, & O. Jokisch. An audiovisual feedback system for acquiring L2 prounication and L2 prosody,

[6] G. G. Lee, H, Lee, J. Song, B. Kim, S. Kang, J. Lee, H. Hwang. "Automatic sentence stress feedback for non-native English learners", Computer Speech and Language, 2017, pp. 29-42.

[7] J. Anderson-hsieh, R. Johnson, & K. Koehler. "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure", Language learning, 1992, pp. 529-555.

[8] J. D. O'Connor, G. Frederick, Intonation of Colloquial English. New Jersey: Prentice Hall Press, 1973.

[9] J. Liu, "On the interlanguage-based teachinig of Japanese segment pronunication: An experiemental phonetic approach", Journal of Japanese Language Study and Research, 2018, pp. 77-84.

[10] M. Zhuang, Y, Bu. "The study of Better Accent Tutor & Visualized Suprasegmental Teaching", Computer-Assisted Foreign Languge Education, 2011, pp. 31-38.

[11] N. Minematsu, Y. Tomiyama, K. Yoshimoto, K. Shimizu, S. Nakagawa, M. Dantsuji, S. Makino. "Development of English speech database read by Japanese to support CALL research", In Proceeding of 2004 ICA, pp. 790-793.

[12] P. Shao, Y. Jia, & A. Li. "Intonation patterns of tag questions for Chinese EFL learners from Shandong dialect", In Proceedings of 12th NCMMSC, 2013, pp. 209-214.

[13] W. F. Katz, & S. Mehta. Visual feedback of tongue movement for novel speech sound learning, Frontiers in Human Neuroscience, 2015, pp. 1-13.

[14] X. Ji, Acquistion of Intonation by Chinese EFL learners——an Empirical Study Based on Experimental Phonetics, Hangzhou: Zhejiang University, 2010.

[15] X. Wang. Phonetic research on English Prosody Acquisition of Chinese Learners based on a Large Comparative Speech Corpus, Beijing: Graduate School of Chinese Academy of Social Sciences, 2010.

[16] X. Zhou, J. Zhang. "A study of intonation patterns of objective clauses for Chinese EFL learners with Liaoning dialect", Comtemporary Foreign Language Education, 2019, pp. 81-92.

[17] Y. Jia, A. Li, C. Tseng. "Construction of English learners from Chinese dialectal regions", Chinese Journal of Phonetics, 2013, pp. 38-45.

[18] Z. Xiong. https://mp.weixin.qq.com/s/apjADRgQj1SREycqPnA5Kw, 2019.

[This paper was published at OCOCOSDA 2021]