

Effect of Music Training on the Production of English Lexical Stress by Chinese English Learners

Hui Feng

School of Foreign Languages and
Literature,
Tianjin University,
Tianjin, China
fenghui@tju.edu.cn

Jie Lian*

College of Foreign Languages,
Taiyuan University of
Technology,
Taiyuan, China
lianjie0330@163.com

Yingjie Zhao

School of Foreign Languages and
Literature,
Tianjin University,
Tianjin, China
victor_zhao@tju.edu.cn

Abstract—Under the guidance of the Theory of Multiple Intelligences, this study aims to find whether music training can improve English stress production among Chinese English learners without music background. Major findings are as follows. (1) In stress production, music training has significant influence on the stress production by Chinese English learners. Specifically, after music training, there has been evident improvement in pitch and intensity in the training group in distinguishing stressed and unstressed syllables in disyllabic pseudowords. Besides, the accuracy of the production of unfamiliar words in the training group has increased by 11.5% on average, compared with that of the control group which has little change. In addition, little effect of music training on duration proportion in stressed syllables is found in this experiment. (2) Chinese English learners' perception of music can be positively transferred to their production of English lexical stress. Such findings provide further proof for the effect of music training on the production of English lexical stress, and propose a method for Chinese English learners to improve their English pronunciation.

Keywords—music training; stress production; second language speech; theory of Multiple Intelligences

I. INTRODUCTION

For many years, English education in China has focused on memorizing words and grammar knowledge on reading and writing levels, with little attention paid to pronunciation and fluency. With the development of globalization, the basic communicative function of language has become increasingly important. Tone and intonation of language thus have become the fundamental elements for better intelligibility in international communication. However, traditional listening and reading imitation has been proved to be unable to improve the accuracy of English speech production properly (Yang, 2010).

Music and language are two methods unique for human beings to express emotions. The relations between these two are becoming major research topics for linguists, psychologists and neurologists in recent years. It is also believed that musical ability affects language ability and many experiments proved that music training has a positive influence on language study (e.g. Besson & Schon, 2001; Ezquerro, 2012; Marques et al., 2007; Mashayekh & Hashemi, 2011; Pei, 2010; Slevc & Miyake, 2006; Zhang & Chang, 2015). Thus, large potential to study the specific

effect of music training on foreign language pronunciation learning in order to find an effective way to make the best of musical elements in second language speech teaching to improve intonation and pronunciation acquisition.

Proposed by Gardner (1983), the Theory of Multiple Intelligences argues that intelligence is multiple which accounts for a group of abilities instead of one ability, and these abilities interact with each other. In addition, every kind of intelligence is cultivable, creative and practical. Gardner's theory is further tested by Patel (2006, 2008) on the neurological level that music and language share one processing mechanism and neural resources in human brain.

Gardner's theory takes into account the influence of individual difference on teaching, and emphasizes the comprehensive development of all kinds of intelligences and sufficient display of individual talents. Therefore, under the Theory of Multiple Intelligences and Patel's further test, this paper mainly discusses whether music training has a positive effect on the production of English lexical stress for Chinese English learners. The study focuses on investigating the effects of music training on the realization of stressed syllables in English lexical words.

This study puts forward two research questions: 1) Can music training improve Chinese English learners' performance in stress production before and after music training? 2) In which way can Chinese English learners improve their stress production? The detailed hypotheses are as follows.

Hypothesis 1: After music training, there is significant improvement in the production of English lexical stress by Chinese English learners.

Hypothesis 2: English lexical stress produced by Chinese English learners are realized by increasing pitch and intensity and extending duration of the stressed syllables.

II. METHODS

A. Participants

In this study, 20 participants (gender balanced) in the training group (PT) and 10 participants (gender balanced) in the control group (PC) are invited to the production experiments. All participants are from North China with standard Chinese (Mandarin) as their language for daily communication. They all know English phonetic symbols and they have no discernable accent in Chinese. None participant has ever learned to play any musical instrument

* Corresponding author

or been trained in singing or ever received any formal music training for at least three months, but they all have the ability to distinguish musical melody and rhythm. All participants have strong motivation to improve their English pronunciation, and are willing to follow the instructions at each stage of the experiment. They are all in their junior year or above, which means they have no compulsive English class to attend. They have not received any English pronunciation training course in the past 6 months and are not attending any such course while they participate in this experiment.

B. Stimuli

Reading materials are composed of three parts: pre-training recording material, training material and post-training recording material. In the selection of recording materials, the following factors are taken into consideration: 1) words in different syllable patterns; 2) familiarity of English words. In this study, 60 words (including disyllabic words and polysyllabic words) are selected randomly from GRE (Graduate Record Examination) wordlists that none of the participants are familiar with. These 60 unfamiliar GRE words are divided evenly and randomly into pre-training and post-training sessions, with stressed syllables and phonetic symbols marked out.

Pseudowords are created on the basis of phonotactic restrictions in a target language (Glaze, 1928), with no real meaning in the lexicon which can not only avoid familiarity of previously known pronunciation, but also imposes strict control on syllabic structures. In this experiment, pseudowords are designed in disyllabic structures “V.CV” and “CV.CV”, with voiced alveolar plosive [d] and voiceless alveolar plosive [t] and five fundamental vowels [a][i][u][e][o]. Since “e” in the words ended with “de” are not pronounced in English, “ade, ede, ide, ode, ude, tade, tede, tide, tode, tude” are removed. In total, there are 40 tokens of pseudowords. Later, 40 pseudowords are doubled to 80 tokens with stressed syllable marked in capitalized letters, such as “Ada /ˈʌdʌ/” and “aDA /ʌˈdʌ/”.

C. Procedure

Before and after music training, all participants in PT and PC are invited to record 30 unfamiliar words in random order. Three English native speakers are invited to assess whether the stress of each word is correctly produced. The mean accuracy scores are used to evaluate the stress production of participants before and after the music training.

After pre-training recording, PT are given weekly music training and recorded randomly after training, including reading pseudowords independently and reading pseudowords in a fixed sentence structure. The music training lasts for 8 weeks including the pitch and loudness perception of musical notes produced from piano, violin, flute and triangle. As for PC, 80 pseudowords tokens are only recorded in the first and last week of music training.

Two months after the recording and production experiments, 8 participants in PT are invited to a follow-up experiment to read 80 pseudowords and 30 unfamiliar words again to test the long-term effect of music training on English stress production. In the follow-up experiment, recordings are scored by three English native speakers to assess the stress placement.

D. Data Extraction and Processing

All sound recordings are first automatically annotated, then syllable boundaries of each word are modified manually in language analysis software Praat. In this study, fundamental frequency (F0), intensity (INT) and duration (DUR) of each target word are extracted; the units of pitch, intensity and duration are hertz (Hz), decibel (dB) and millisecond (ms) respectively.

For disyllabic pseudowords, $\Delta F0$, ΔINT and $\%DUR$ are calculated to measure the acoustic differences between stressed and unstressed syllables:

$$\Delta F0_i = F0_i - f0_i \quad (1)$$

$$\Delta INT_i = INT_i - int_i \quad (2)$$

$$\%DUR_i = \frac{DUR_i}{Dur_w} \times 100\% \quad (3)$$

where $F0_i$, INT_i , DUR_i are F0, intensity and duration of the stressed syllables; $f0_i$, int_i are F0 and intensity of the unstressed syllables; Dur_w is the entire duration of the target word. $\Delta F0$ is the difference between F0 in stressed syllable and F0 in unstressed one in a disyllabic pseudoword; ΔINT is the difference between intensity in stressed syllable and intensity in unstressed one in a disyllabic pseudoword; $\%DUR$ is the proportion of the length of stressed syllable in the whole word length.

III. RESULTS AND DISCUSSION

In stress perception, small variation is found in the accuracy of English stress perception after music training. Before music training, all participants' ability in English lexical stress perception is generally at a high level (accuracy around 90-95%) with no significant individual differences, while after music training, stress perception accuracy in the training group is still around 90-95% with little variation.

In this study, participants' stress production data are collected by recording pseudowords and lexical words. Before music training, it is noticed that 7 participants in the training group have difficulties in stress realization on the condition that they know where the stressed syllable should be, but they do not know how to realize the intended stress. They tend to produce higher pitch in the first syllable, stronger intensity and longer duration in the second syllable, which makes it even hard for listeners to identify stress placement. Therefore, in this part, the 20 participants in the training group are divided into Group A (7 participants with difficulty in stress realization) and Group B (13 participants with little difficulty in stress realization).

A. Production of Pseudowords

Generally, in disyllabic words, the stressed syllables are pronounced with higher frequency, stronger intensity and longer duration compared with the unstressed syllables (Fry, 1958). As for the production of F0 and intensity in pseudowords in this study, $\Delta F0$ and ΔINT are the main parameters to indicate the prominence of pitch and intensity of the stressed syllables in disyllabic tokens. The mean $\Delta F0$ and ΔINT changes of Group A and Group B in PT before and after music training are shown in Fig. 1, data are analyzed in t-test.

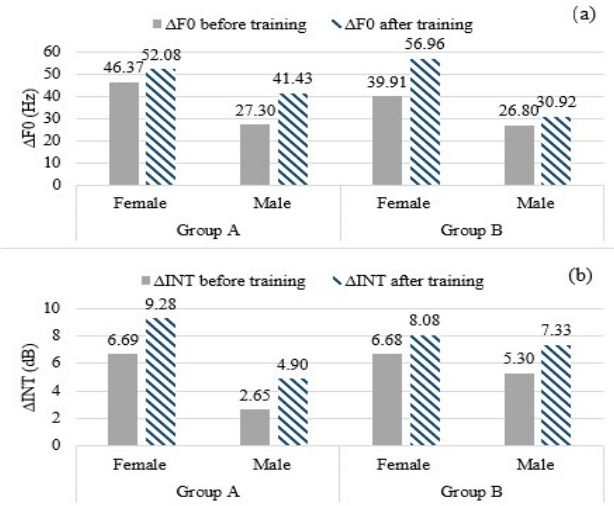


Figure 1. Mean $\Delta F0$ and ΔINT changes of PT before and after music training

Fig. 1 reveals that, overall, there is significant improvement in $\Delta F0$ ($t=3.25$, $P=.047$) and ΔINT ($t=8.27$, $P=.003$) in disyllabic pseudowords produced by the two groups of PT after music training. Female participants in Group B make greater progress in $\Delta F0$ than that of female participants in Group A (17.05Hz vs. 5.71Hz), while male participants in Group A performed better in F0 production than male participants in Group B (14.1Hz vs. 4.1Hz). Besides, participants in Group A have greater enhancement in ΔINT than that of participants in Group B (2.58dB vs. 1.40dB for female, 2.25dB vs. 2.03dB for male). That is to say, in intensity production, the improvement of participants in Group A is more significant.

Z-score normalization (standard score) method is utilized in this part to eliminate the effect of gender differences in F0 and intensity. The normalized mean $\Delta F0$ and ΔINT , and %DUR of tokens with word-initial stress and tokens with word-final stress produced by PT and PC before and after music training are listed in Table I.

Results in Table I indicate that for participants in the training group, there is significant improvement ($t=4.34$, $P=.022$) in $\Delta F0$ and ΔINT of disyllabic tokens after music training. On average, PT have improved by 22.03% in $\Delta F0$ and 79.67% in ΔINT in stressed syllables in tokens with word-initial stress; they have improved by 60.44% in $\Delta F0$ and 16.59% in ΔINT in stressed syllables in tokens with word-final stress. However, for the control group, there is little change ($t=1.84$, $P=.161$) in $\Delta F0$ and ΔINT of disyllabic tokens. Such findings suggest that there is significant enhancement in the prominence of pitch and intensity of the stressed syllables produced by the training group after they received music training.

As for duration production, on average, before music training, all participants put less than half of time (PT 47.34%, PC 46.16%) on the stressed syllables when they pronounce tokens with word-initial stress, while they produce much longer (PT 69.52%, PC 68.77%) stress in tokens with word-final stress which may resulted from the phenomenon that final-stressed syllables will extend duration to realize focus. The proportion results show little changes in duration proportion of stressed syllables after music training in both groups. Such findings suggest that

music training has little influence on the duration production of stressed syllables in disyllabic tokens.

TABLE I. $\Delta F0$, ΔINT and %DUR of disyllabic tokens production

		Initial stress			Final stress		
		before	after	rate(%)	before	after	rate(%)
PT	$\Delta F0$	1.38	1.69	22.03	0.48	0.77	60.44
	ΔINT	0.39	0.71	79.67	0.30	0.35	16.59
	%DUR	47.34	48.46	2.37	69.52	68.23	-1.86
PC	$\Delta F0$	1.75	1.75	0.16	1.57	1.52	-1.11
	ΔINT	0.83	0.84	0.44	0.67	0.66	-1.50
	%DUR	46.16	45.78	-0.82	68.77	69.05	0.41

Notes: here, $rate=100\%*(after-before)/before$.

In the follow-up experiment, 8 participants still maintain the improvement of pitch and intensity in stress production of disyllabic tokens as they did at the end of the music training (average improvement 40.4% vs. 44.7% without significant statistical difference, $t=1.872$, $P=0.473$).

To sum up, compared with the control group which receives no music training, there is significant improvement in stress prominence of F0 and intensity between stressed and unstressed syllables in disyllabic pseudowords produced in the training group after music training. Little effect of music training on duration proportion in stressed syllables is discovered in this experiment.

B. Production of Lexical Words

Participants in both the training group and the control group are invited to read 30 unfamiliar words with stress and phonetic symbols marked out before music training, and read another group of 30 unfamiliar words after training. Audio recordings are sent to three English native speakers to assess whether the stress placement is correct.

The accuracy of different groups in PT before and after music training is calculated and presented in Fig. 2. It can be indicated from Fig. 2 that overall accuracy of unfamiliar words has increased by 11.5% (=88.83%-77.33%) which means there is overall enhancement in PT's accuracy of reading unfamiliar words after music training (88.83% vs. 77.33%; $t=6.97$, $P=.000$) with statistical significance. Group B performs better than Group A in reading unfamiliar words (82.82% vs. 67.14% before training, 91.03% vs. 84.76% after training) with no statistical significance, while Group A makes greater progress in words accuracy (17.62% vs. 8.20%).

Accuracy (numbers of the words with correct stress pronunciation) in pre-training and post-training records of the training group and the control group are analyzed in SPSS for paired-samples t-test. The results of t-test show that, there exist significant differences in the training group in the accuracy of reading unfamiliar words before and after training ($t=2.09$, $P=.000$), while in the control group, the difference in the accuracy of reading unfamiliar words is not remarkable ($t=2.26$, $P=.343$). Such findings indicate that PT's accuracy in reading unfamiliar words have improved after music training.

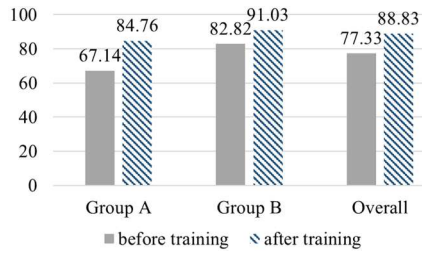


Figure 2. Mean accuracy (%) of reading unfamiliar words by PT

In the follow-up experiment, eight PT's recordings are scored by three English native speakers to identify the stress placement. The accuracy in the follow-up experiment is still over 82%.

C. Effect of Music Perception on Stress Production

After musical pitch and loudness distinguishing training, PT produce stress by increasing $\Delta F0$ and ΔINT between the stressed and unstressed syllables, which means that there is significant improvement in participants' stress production in disyllabic pseudowords after music training. In this study, the mean accuracy improvement in musical pitch and loudness identification before and after music training is reckoned as the music perception ability of PT. The rate of $\Delta F0$ and ΔINT in disyllabic pseudowords for each participant (variations between mean $\Delta F0$ and ΔINT before training, and mean $\Delta F0$ and ΔINT after training) is defined as variation of the improvement of speech production. Variation values of accuracy of musical pitch and loudness as well as $\Delta F0$ and ΔINT in disyllabic pseudowords are normalized to eliminate random differences. Fig. 3 presents the correlation between perception of music and production of pseudowords.

As is shown in Fig. 3(a), variation value of pitch perception of music is basically in positive linear correlation with variation value of $F0$ production of pseudowords ($r=.874$, high correlation). In other words, most PT with higher perception in musical pitch produce greater $\Delta F0$ in stressed syllables of disyllabic pseudowords, which indicates that in terms of pitch, the improvement in music perception can be mostly transferred to stress realization in speech production. Fig. 3(b) indicates that, variation value of loudness perception of music is basically in positive linear correlation with variation value of intensity production of pseudowords ($r=.746$, moderate correlation). That is to say, the majority of PT with higher perception in musical loudness produce larger ΔINT in stressed syllables of disyllabic pseudowords, which indicates that in terms of loudness, the improvement in music perception can be largely transferred to the stress realization in speech production.

Therefore, it can be summarized that after music training, PT with greater progress in musical pitch and loudness perception could produce stress in a more significant way by increasing $F0$ and intensity in the stressed syllables. As for PT with little improvement in music perception, their enhancement in the production of stressed syllables is not quite obvious. In other words, PT with higher sensitivity in music are more likely to benefit from music training to improve their stress production in speech.

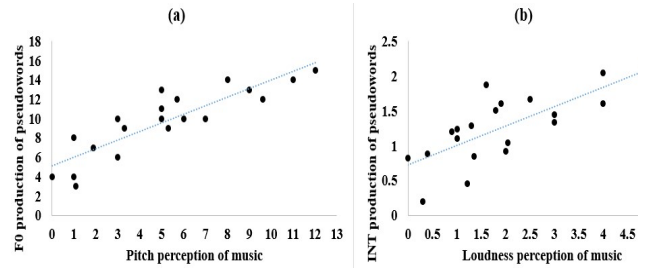


Figure 3. Correlation between perception of music and production of pseudowords

IV. CONCLUSION

Based on Gardner's Theory of Multiple Intelligences, this study aims to find whether Chinese English learners without music background can improve their production of English stress after a set of music training schedule by conducting phonetic experiments. Major findings are as follows. (1) In stress production, there is significant influence of music training on the stress production by Chinese English learners. After two-month music training, the training group make great progress in distinguishing stressed and unstressed syllables in disyllabic pseudowords. They mainly achieve it by increasing pitch (around 42%) and intensity (around 87%) in the stressed syllables. Such improvement is not found in the control group. Besides, the accuracy of the production of unfamiliar words in the training group has increased by 11.5% on average, compared with the control group which has little change. In addition, the proportion results show small variation in duration proportion of the stressed syllables after music training in both PT and PC, which means that little effect of music training on duration proportion in stressed syllables is discovered in this experiment. (2) Chinese English learners' perception of music can be positively transferred to the production of English lexical stress. After music training, participants in the training group with greater progress in music perception can produce stressed syllables in a more significant way by increasing pitch and intensity.

Such findings provide further proof for the effect of music training on the production of English lexical stress, and propose a learning method for Chinese English learners to improve their English pronunciation. The set of music training which has been proved effective can be applied to language pronunciation classes, and teachers can use these musical methods to improve the efficiency of language teaching.

ACKNOWLEDGMENTS

The research is supported partially by the Major Program of National Social Science Foundation of China (No. 15ZDB103), and the Social Science Foundation of Tianjin, China (No. TJWW17-010 and TJWW19-009).

REFERENCES

- [1] J. Yang, "A study of the effects of repeated listening and imitation on the production of English stress patterns," *Foreign Languages Research*, vol. 4, pp. 9-16, April 2010.
- [2] M. Besson, D. Schon, "Comparison between language and

- music,” *Annals New York Academy of Sciences*, vol. 930, pp. 232-258, 2001.
- [3] A. O. Ezquerro, “Exploring the relationship between musical aptitude and degree of comprehensibility in L2 speech: Assessment of a group of English learners of Spanish in the English secondary school system,” *Porta Linguarum*, vol. 17, pp. 223-236, 2012.
 - [4] C. Marques, S. Moreno, S. L. Castro, M. Besson, “Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence,” *Journal of Cognitive Neuroscience*, vol. 19, pp. 1453-1463, 2007.
 - [5] M. Mashayekh, M. Hashemi, “The impacts of music on language learners’ performance,” *Procedia - Social and Behavioral Sciences*, vol. 30, pp. 2186-2190, 2011.
 - [6] Z. Pei, *An Empirical Study on the Effects of Musical Aptitude on EFL Phonological Learning*. Nanjing: Nanjing University, 2010.
 - [7] R. Slevc, A. Miyake, “Individual differences in second-language proficiency: Does musical ability matter?” *Psychological Science*, vol. 17, pp. 675-681, 2006.
 - [8] H. Gardner. *Frames of Mind: The Theory of Multiple Intelligence*. New York: New York Basic Books, 1983.
 - [9] A. D. Patel, “Musical rhythm, linguistic rhythm, and human evolution,” *Music Perception*, vol. 24, pp. 99-104, 2006.
 - [10] A. D. Patel, *Music, Language, and the Brain*. Oxford: Oxford University Press, 2008.
 - [11] J. A. Glaze, “The association value of non-sense syllables,” *Pedagogical Seminary and Journal of Genetic Psychology*, vol. 35, pp. 255-269, 1928.
 - [12] D. Fry, “Experiments in the perception of stress,” *Language and Speech*, vol. 1, pp. 120-152, 1958.
 - [13] H. Zhang, H. Chang, “Effects of musical training on Chinese learners’ perception of English suprasegmental,” *Second Language Learning Research*, vol. 1, pp. 73-83, 2015.