Effects of Music Training Experience on the Production of English Rhythm by Chinese Learners

Chenyu Li, Ying Chen*, Ruizhe Wang and Yujia Zhang
Nanjing University of Science and Technology, China
E-mail: ychen@njust.edu.cn Tel: +86-25-84315885

Abstract

Rhythm is a pivotal element of speech prosody, playing a crucial role in intelligibility and presenting considerable challenges for second language (L2) learners. Previous studies have demonstrated that Chinese learners of English as a foreign language (EFL) encounter difficulties in mastering the duration of vowels and consonants, as well as in adjusting pitch patterns when producing English rhythms. This study aims to explore the effect of music training on the production of English rhythm by Chinese EFL learners. The results, based on comparisons of various rhythmic parameters, show significant differences in VarcoV and VarcoC between the music and non-music groups among low-proficiency learners. This finding suggests that music training facilitates improving the English rhythm production among Chinese EFL learners.

Index Terms: Speech rhythm, Music training, EFL, e-OPREA

I. INTRODUCTION

Researchers have classified languages into the rhythmic categories of syllable-timed, stress-timed, and mora-timed. English has been traditionally categorized as a stress-timed language, mainly due to its intricate syllable structure, vowel reduction, and fluctuating stress patterns [13]. Mandarin Chinese, on the other hand, has been classified as a syllable-timed language, with acoustic evidence supporting this classification [3]. Although the strict classification of languages into rhythmic categories remains debatable [5], researchers have developed various parameters to quantify rhythmic characteristics across languages [7] [8] [9] [10]. The foundational work by Ramus et al. [7] introduced interval measures including %V, ΔV, and ΔC, which were later refined by Dellwo[10] through speech rate normalization (VarcoV and VarcoC). Another significant development was the introduction of PVI (Pairwise Variability Indices) by Low et al. [8] and Grabe and Low [9], which measure durational contrast between successive elements.

As a pivotal element of speech prosody, rhythm plays an essential role in assessing speech production of non-native speakers and significantly affects the comprehensibility of speech[1][2]. As a suprasegmental feature, rhythm presents considerable acquisition challenges for second language (L2) learners. The substantial differences in rhythm contexts between the two languages may result in difficulties in acquiring English rhythm for Chinese EFL learners [4]. Studies examining English rhythm production by Chinese EFL learners have revealed a stronger tendency towards isochrony while their rhythmic patterns show relatively minor deviations from native speakers [11] [12] [13] [14]. However, these studies have been limited by their focus on sentence-level rather than discourse-level analysis [14] [15] and their predominant emphasis on duration-based indices while overlooking pitch-based measurements [12] [15].

The intrinsic connection between music and language is evidenced by their shared acoustic properties, including pitch, duration, intensity, and prosodic rhythm [16]. Building on these inherent similarities, recent advances in cognitive neuroscience have established theoretical frameworks explaining how music training enhances L2 speech acquisition. The expanded OPERA hypothesis (e-OPREA) [17] proposes that music and speech share neural processing mechanisms in the brain, with music training enhancing these shared mechanisms through five key components: (1) **O**verlap in brain networks for processing speech and music, (2) **P**recision requirements (higher in music), (3) **E**motional reward, (4) **R**epetition, and (5) focused **A**ttention. This theoretical framework is further strengthened by the concept of domain-general sharpening, which suggests that music training refines the perceptual encoding of auditory features shared by both language and music through cortical-subcortical circuits.

Empirical studies have provided substantial evidence supporting these theoretical predictions. Early investigations by Nakata [21] and Tanaka & Nakamura [22] established a fundamental correlation between music ability and L2 pronunciation. Subsequent research has demonstrated that music training specifically influences learners' perception and production of L2 suprasegmental features [18] [19] [20]. For instance, Llanes-Coromina et al. [23] found that music rhythm training particularly enhanced L2 pronunciation fluency and comprehensibility. In the context of Chinese students learning English, Pei et al. [24] demonstrated that Chinese EFL learners with music training exhibited superior performance in suprasegmental aspects of L2 production. This effect was observed across learners with different native languages, including French, German, Russian, and Japanese.

On the basis of the theoretical correlation between music and speech and the research gaps in the studies of L2 rhythm, the following research questions were proposed:

(1) Does music training experience affect the English rhythm produced by Chinese EFL learners? If so, on which specific rhythmic parameters is this effect more clearly observed than others?

(2) In which type of context, sentence level or discourse level, is English rhythm affected more saliently by the Chinese EFL learners' music training experience?

(3) How does English proficiency level interact with music training background as the effects of Chinese EFL learners' production of English rhythm?

II. METHODS

A. Participants

Sixty-one native Mandarin speakers (Mean = 20.79 yrs., SD = 2.56, 31F/30M) from Nanjing University of Science and Technology and Nanjing Normal University participated in the study. All were non-English majors with minimum nine years of English learning and no hearing or speech disorders. Participants were grouped by English proficiency (CET-4 scores 425-500 for low proficiency; CET-6 scores above 550 for high proficiency) and music training (non-music: only school music education; music: minimum six years of vocal/instrumental training plus passing music ear test [41] via NAODAO [42]).

B. Stimuli

Six sentence types with different intonational patterns in American English were selected as the sentence stimuli from Teschner and Whitley [25]:

Yes-No question, e.g., *Do you want to buy some new clothes?*

Statement, e.g., *He answers questions actively in class.*

Enumeration, *e.g., They received pens, paper and glue.*

Selective question, e.g., *Do you want to take a taxi or a bus?*

Tag, e.g., *My mother will take me to the park if she is available.*

Complex sentence, e.g., *My mother wouldn't let me go because I didn't finish my homework.*

In total, 42 stimulus sentences were selected in terms of sentence type, yielding 7 blocks (i.e., Blocks A to G) with each block possessing six intonational patterns, including falling, rising, flat, rise-fall, fall-rise, and low-rising, reflecting typical English intonation contours [25].

The discourse stimuli were selected from the Speech Accent Archive, an authoritative database of language accents from various language backgrounds [26]:

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

C. Recording

Recording was conducted in the soundproof booth at LIPA Lab, Nanjing University of Science and Technology, across two days to avoid practice and fatigue effects. Participants read aloud the sentence stimuli in a random order presented via PsychoPy and a passage five times at normal speech rate. Each participant produced 42 sentences and one passage and recorded by a Marantz PMD661 recorder and a Shure SM10A-CN microphone with 44,100 Hz sampling rate. The self-paced experiments lasted approximately 30 minutes per participant.

D. Acoustic parameters

The acoustic analysis focused on two main parameters: Variation Coefficient (Varco) and Pairwise Variability Index (PVI). Varco measures global variation in duration and pitch, including VarcoV for vowel duration, VarcoC for consonant duration, and VarcoV(F0ex) for vowel pitch, while PVI calculates local variation between adjacent segments, including nPVI-V for vowel duration, rPVI-C for consonant duration, and nPVI-V(F0ex) for vowel pitch.

VarcoV, VarcoC

$$\mathrm{Varco(x)} = 100 \times \left(\frac{\Delta \mathrm{d_x}}{\mathrm{d_x}}\right) \quad (1)$$

Where x represents the phonetic unit, Δd_x is the standard deviation of the duration of the phonetic unit, and d_x is the average duration of the phonetic unit.

nPVI-V, rPVI-C

$$\mathrm{nPVI} = 100 \times \left(\frac{\sum_{k=1}^{m-1} |d_k - d_{k+1}|}{\frac{d_k + d_{k+1}}{2}}\right) \div (m-1) \quad (2)$$

$$\mathrm{rPVI} = \left(\sum_{k=1}^{m} |d_k - d_{k+1}|\right) \div (m-1) \quad (3)$$

Where m is the number of a certain phonetic unit and d_k is the duration of the phonetic unit number k.

VarcoV(F0ex)、nPVI-V (F0ex)

$$\mathrm{Varco\ V(FO_{ex})} = 100 \times \frac{\Delta\ \mathrm{FO_{ex}}}{\mathrm{FO_{ex}}} \quad (4)$$

Where FO_{ex} is the fundamental frequency excursion($F0_{\max} - F0_{\min}$) of a phonetic unit, ΔFO_{ex} is the standard deviation of the fundamental frequency excursion and $\overline{FO}_{ex}$ is the mean value of the fundamental frequency excursion.

$$\mathrm{nPVI - V\ (FO_{ex})} = 100 \times \left(\frac{\sum_{k=1}^{m-1} |\mathrm{FO_{ex_k}} - \mathrm{FO_{ex_{k+1}}}|}{\frac{\mathrm{FO_{ex_k}} + \mathrm{FO_{ex_{k+1}}}}{2}}\right) \div (m-1) \quad (5)$$

Where m is the number of phonetic units and FO_{ex_k} is the fundamental frequency excursion of the k-th phonetic unit.

E. Data analysis

Acoustic data were extracted using Praat [28] and the ProsodyPro [29] script. The measurements included duration (ms) and F0ex (semitone) from consonant intervals (end of preceding consonant/pause to beginning of next consonant) and vowel intervals (end of vowel/pause to start of next vowel).

Statistical analyses were conducted using R [30] and the lmerTest [31] package for linear mixed-effects models. Dependent variables included VarcoV, VarcoC, nPVI-V, rPVI-C, VarcoV(F0ex), and nPVI-V(F0ex). Fixed factors were music experience (music/non-music training), English proficiency (high/low), and rhythm context (sentence/discourse level), with stimuli ID and speaker ID as random factors. Post-hoc comparisons and visualizations were

performed using the emmeans [32] and ggplot2 [33] packages. All statistical contrasts were coded as Non-music vs. Music Group, where negative beta coefficients indicate higher values in the music training group.

III. RESULTS

A. *Timing patterns*

Variation Coefficient (Varco C, Varco V)

Statistical analyses revealed significant main effects of music training background ($F(1, 316.6) = 4.8$, $p < 0.05$), English proficiency ($F(1, 316.6) = 4.5$, $p < 0.05$), and context of rhythm ($F(1, 42.5) = 63.7$, $p = 0.001$) on VarcoC. Significant interactions were found between music background and English proficiency ($F(1, 316.6) = 1.4$, $p < 0.01$), and between English proficiency and context of rhythm ($F(1, 6632.5) = 3.9$, $p < 0.05$)

Post-hoc analyses showed significant differences in VarcoC between proficiency levels at discourse level ($\beta = 9.92$, $SE = 4.52$, $t = 2.19$, $p < 0.01$) but not at sentence level ($\beta = 1.35$, $SE = 1.65$, $t = 0.82$, $p = 0.414$). Among low-proficiency learners, the music group showed significantly higher VarcoC values ($\beta = -15.98$, $SE = 5.75$, $t = -2.781$, $p = 0.0054$), while no significant difference was found for high-proficiency learners ($\beta = -2.69$, $SE = 4.06$, $t = -0.66$, $p = 0.507$).

For VarcoV, significant main effects were found for both music experience ($F(1, 144.2) = 5.38$, $p < 0.05$) and English proficiency ($F(1, 144.2) = 10.71$, $p < 0.01$), with a significant interaction between them ($F(1, 6630.2) = 4.46$, $p < 0.05$). Post-hoc tests revealed significantly higher VarcoV values in the music group among low-proficiency learners at sentence level ($\beta = -3.71$, $SE = 1.49$, $t = -2.489$, $p = 0.0128$), but not among high-proficiency learners ($\beta = -3.26$, $SE = 2.32$, $t = -1.40$, $p = 0.161$).

Figure 1 illustrates these rhythm parameters (VarcoV and VarcoC), clearly showing the significant differences between the music and non-music groups, particularly for low-proficiency learners.

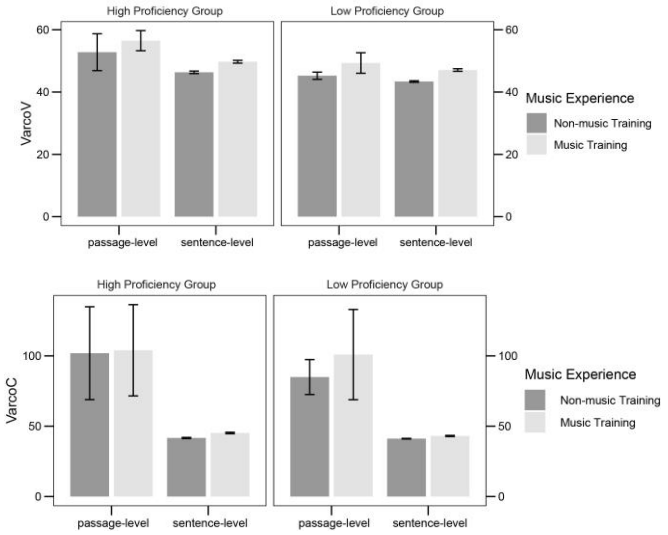

Figure 1 Bar charts of the rhythm parameters VarcoV and VarcoC

Pairwise Variability Index (nPVI-V, rPVI-C)

For Pairwise Variability Indices (nPVI-V and rPVI-C), no significant main effects were found for music experience (nPVI-V: $F(1, 146.6) = 0.79$, $p = 0.37$; rPVI-C: $F(1, 83) = 0.25$, $p = 0.61$), English proficiency (nPVI-V: $F(1, 146.6) = 0.001$, $p = 0.97$; rPVI-C: $F(1, 83) = 0.64$, $p = 0.42$), or context of rhythm (nPVI-V: $F(1, 39) = 0.0003$, $p = 0.98$; rPVI-C: $F(1, 41.7) = 0.22$, $p = 0.63$). Similarly, no significant two-way or three-way interactions were observed among these factors (all $ps > 0.38$).

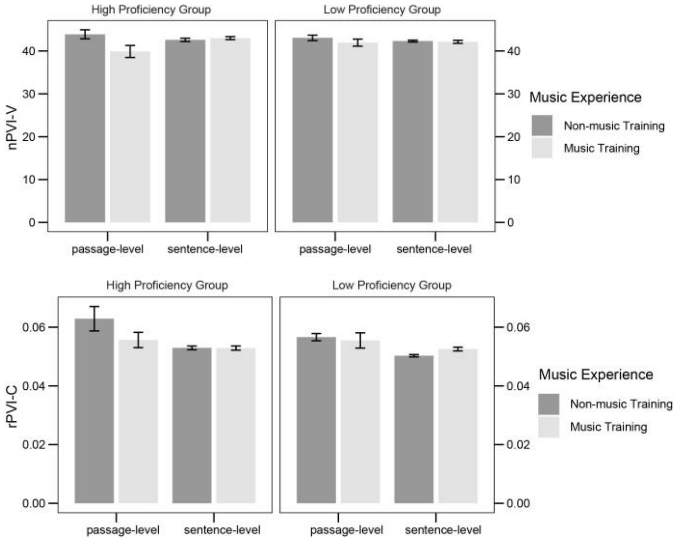

Figure 2 Bar charts of the rhythm parameters nPVI-V and rPVI-C

B. *Pitch patterns*

Statistical analyses revealed a significant interaction between English proficiency and music experience for both rhythmic parameters Varco(F0ex) and nPVI(F0ex) ($F(1, 320.7) = 4.04$, $p < 0.05$). No significant two-way interactions were found between music experience and context of rhythm ($F(1, 6633.8) = 0.28$, $p = 0.59$), or English proficiency and context of rhythm ($F(1, 6633.8) = 0.78$, $p = 0.37$). The three-way interaction among these factors showed no significance ($F(1, 6633.8) = 0.21$, $p = 0.88$).

For Varco(F0ex) and nPVI(F0ex), Tukey HSD post-hoc tests revealed no significant differences between music and non-music groups across proficiency levels. Among low-proficiency learners, the differences were ($\beta = 6.57$, $SE = 7.95$, $t = 0.826$, p = 0.4088) for Varco(F0ex) and ($\beta = 5.90$, $SE = 3.63$, $t = 1.62$, $p = 0.104$) for nPVI(F0ex). For high-proficiency learners, the values were ($\beta = -12.42$, $SE = 9.67$, $t = -1.284$, $p = 0.1992$) and ($\beta = -6.15$, $SE = 4.40$, $t = -1.40$, $p = 0.163$) respectively.

As shown in Fig. 3, both rhythmic parameters at the discourse level were higher in the music group among high-proficiency Chinese EFL learners, though these differences were not statistically significant.

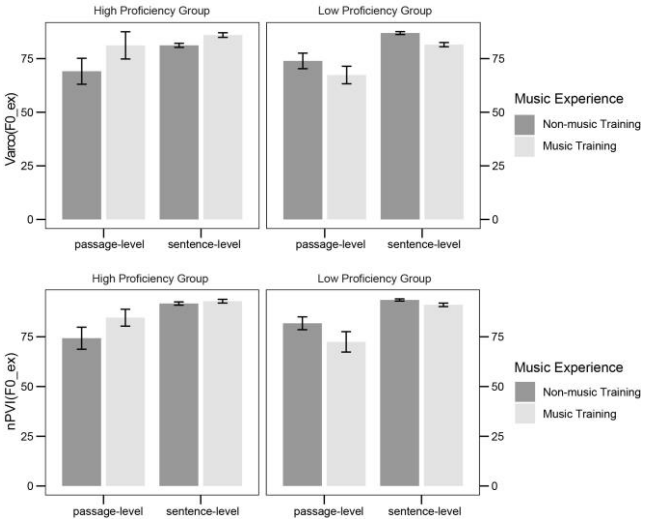

Figure 3 Bar charts of the rhythm parameters VarcoV(F0ex) and nPVI-V(F0ex)

IV. DISCUSSION

The current study explored the impact of music training on English rhythm production among Chinese EFL learners, revealing two major findings. First, music training significantly enhanced rhythm production among low-proficiency learners, as indicated by higher VarcoV and VarcoC values. Second, this enhancement varied with proficiency levels, with stronger benefits for low-proficiency learners. These findings align with prior research on Mandarin speakers' typically lower rhythmic variability in L2 English, suggesting that music training helps low-proficiency learners better process rhythmic cues and enhance their English rhythm production.

A. The Effect of Music Training

Previous studies [11] [12] [14] [35] have consistently demonstrated that Chinese learners of English with L1 Mandarin exhibited lower variability in vocalic and consonantal durations than native speakers of English, resulting in lower values of VarcoV and VarcoC. This pattern reflects the fundamental rhythmic differences between Mandarin (a syllable-timed language) and English (a stress-timed language), where Mandarin speakers tend to produce more isochronous speech with less salient durational contrasts [13]. Our findings suggest that music training may facilitate overcoming this L1 interference by enhancing learners' sensitivity to durational contrasts, particularly benefiting low-proficiency learners who are still establishing new rhythmic patterns. The present study confirms that music training significantly enhances VarcoV and VarcoC for low-proficiency learners, increasing their ability to produce rhythmically varied speech. Future research comparing these learners' performance with native speakers will provide a clearer understanding of the extent of these improvements. This result supports the notion that Mandarin learners of English often struggle with vowel and consonant duration variability [36], which can be mitigated through music training.

For high-proficiency learners, the difference between the music and non-music groups was not statistically significant, although the music group showed a slight advantage. This suggests that while music training enhances rhythmic performance at lower proficiency levels, its effect diminishes as proficiency increases. One possible explanation is the "threshold effect," where high-proficiency learners already possess strong rhythmic control, reducing the incremental benefit of music training.

B. Differences in Rhythmic Parameters

Unlike VarcoV and VarcoC, the rhythmic indices nPVI-V and rPVI-C did not show significant main effects or interactions in the linear mixed model. This may be due to the limited variability of these indices between L1 English speakers and Chinese EFL learners. Unlike VarcoV and VarcoC, which directly measure the duration of speech units, nPVI-V and rPVI-C focus on the contrast between adjacent vowel and consonant durations. These contrasts may not be as pronounced in Chinese EFL learners, explaining the non-significant results. Future research should further investigate this by examining different task types or adjusting for the influence of speech rate [36]. The underlying mechanisms remain unclear, and future studies will consider increasing sample sizes and diversifying music training backgrounds to better understand the differential effects of rhythm-focused versus melody-focused training on L2 rhythm acquisition. Additionally, further exploration of task effects and the role of pitch patterns may provide deeper insights into how music training influences L2 speech rhythm.

C. Pitch Patterns and Music Training

The pitch pattern indices VarcoV (F0ex) and nPVI-V (F0ex) exhibited significant interactions between music training and English proficiency. However, post-hoc analysis (Tukey HSD test) did not yield statistically significant results. As shown in Figure 3, high-proficiency learners in the music group showed higher Varco(F0ex) and nPVI(F0ex) values at the discourse level compared to the non-music group, but these differences did not reach statistical significance.

This result could be due to the fact that high-proficiency learners already exhibit a high level of rhythmic variability, making the additional effect of music training less pronounced. Nevertheless, the findings suggest that high-proficiency learners demonstrate greater variability in rhythm production when processing both pitch and timing cues in English rhythm. This partially supports the e-OPERA hypothesis, which posits that music imposes higher cognitive demands than speech [34] [17]. As a result, its effects are more pronounced in learners with lower proficiency.

D. The Cross-Domain Transfer Between Music and Language

The findings of this study, particularly those illustrated in Figures 1 and 3, can be explained through cross-domain transfer mechanisms between music and language. Learners

with music training demonstrated advantages in rhythm production, as reflected in the increased variability of VarcoV and VarcoC in low-proficiency learners (Figure 1) and greater pitch and timing variability in high-proficiency learners (Figure 3).

This result supports the hypothesis that music training enhances phonetic processing and rhythm acquisition by improving the ability to process acoustic signals in both music and linguistic contexts. Long-term music experience enhances the brain capacity to process and adapt to acoustic signals, which in turn benefits language learning by improving prosodic perception and production [37] [38] [39] [40]. Studies suggest that long-term experience in one domain (e.g., music) can enhance acoustic processing in another domain (e.g., language) and influence the formation of abstract representations in other areas [40].

However, this study does not fully clarify the extent to which cross-domain transfer mechanisms impact L2 speech rhythm. Future research will include acoustic data from native speakers of English to further explore how music influences the acquisition of L2 speech rhythm. Additionally, this study adopted a cross-sectional design, demonstrating correlations between music training and English rhythm production rather than causal relationships. Future research will employ randomized controlled trials and longitudinal studies to investigate causal effects and learners' progress over time.

V. Conclusions

The present study provides evidence that music training benefits English rhythm produced by Chinese EFL learners, particularly enhancing rhythmic parameters—VarcoV and VarcoC among low-proficiency learners. These findings suggest that music training can be a potentially effective method for L2 rhythm acquisition, with implications for integrating music elements into EFL teaching. Future research will examine acoustic data of native English speakers for deeper insights into the effects of music training on L2 rhythm acquisition.

VI. Acknowledgements

This work was supported by the grant of National Social Science Foundation of China, approval number 19BYY043, and National Undergraduate Training Program for Innovation and Entrepreneurship of China.

References

[1] C. Gussenhoven, *On the Grammar and Semantics of Sentence Accents*, vol. 16. Berlin, Germany: Walter de Gruyter GmbH & Co KG, 2014.

[2] R. Ellis, "Task-based language teaching: Sorting out the misunderstandings," *Int. J. Appl. Linguist.*, no. 3, pp. 221-246, 2009.

[3] H. Lin and Q. Wang, "Mandarin rhythm: an acoustic study," *J. Chin. Lang. Comput.*, vol. 17, no. 3, pp. 127-140, 2007.

[4] T. Odlin, *Language Transfer*. Cambridge, UK: Cambridge University Press, 1989.

[5] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," *Papers in Laboratory Phonology*, vol. 7, no. 515-546, pp. 1-16, 2002.

[6] L. E. Ling, E. Grabe, and F. Nolan, "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English," *Language and Speech*, vol. 43, no. 4, pp. 377-401, 2000.

[7] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, no. 3, pp. 265-292, 1999.

[8] E. L. Low, E. Grabe, and F. Nolan, "Quantitative characterization of speech rhythm: Syllable-timing in Singapore English," *Language and Speech*, vol. 43, pp. 377-401, 2000.

[9] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," *Lab. Phonol.*, vol. 7, pp. 515-546, 2002.

[10] V. Dellwo and P. Wagner, "Relations between language rhythm and speech rate," in *Proc. 15th Int. Congr. Phonetic Sci.*, Barcelona, 2003.

[11] H. L. Jian, "On the syllable timing of Taiwan English," in *Proc. Speech Prosody 2004*, Nara, Japan, 2004.

[12] J. Yu, Y. Liao, and Y. Lin, "Study on the rhythm characteristics of Chinese English learners under different spoken task types," *Foreign Languages and Their Teaching*, no. 02, pp. 79-90+147-148, 2022, doi: 10.13458/j.cnki.flatt.004848. [in Chinese]

[13] R. Fuchs and E. M. Wunder, "A sonority-based account of speech rhythm in Chinese learners of English," in *Universal or Diverse Paths to English Phonology*, pp. 165-183, 2015.

[14] J. Yang and H. Chen, "Prosody of second language oral output: A literature review related to reading aloud," *Foreign Language Research*, no. 5, pp. 46-50, 2005. [in Chinese]

[15] M. Yang, "Learner chunk 'staccato' phenomenon affecting oral fluency," *Journal of Shenyang University (Social Science Edition)*, no. 2, pp. 229-233, 2019. [in Chinese]

[16] F. A. Russo and M. K. Pichora-Fuller, "Tune in or tune out: Age-related differences in listening to speech in music," *Ear and Hearing*, vol. 29, no. 5, pp. 746-760, 2008.

[17] A. D. Patel, "Can nonlinguistic music training change the way the brain processes speech? The expanded OPERA hypothesis," *Hearing Research*, vol. 308, pp. 98-108, 2014.

[18] C. Marie, F. Delogu, G. Lampis, et al., "Influence of music expertise on segmental and tonal processing in Mandarin Chinese," J. Cogn. Neurosci., vol. 23, no. 10, pp. 2701-2715, 2011.

[19] Z. Pei, Y. Wu, X. Xiang, et al., "The effects of music aptitude and music training on phonological production in foreign languages," English Language Teaching, vol. 9, no. 6, pp. 19-29, 2016.

[20] N. Perrachione, J. Lee, L. Ha, et al., "Learning a novel phonological contrast depends on interactions between

individual differences and training paradigm design," J. Acoust. Soc. Am., vol. 130, no. 1, pp. 461-472, 2011.

[21] H. Nakata, "Correlations between music and Japanese phonetic aptitudes by native speakers of English," *Reading Working Papers in Linguistics*, no. 6, pp. 1-23, 2002.

[22] A. Tanaka and K. Nakamura, "Auditory memory and proficiency of second language speaking: A latent variable analysis approach," *Psychol. Rep.*, vol. 95, pp. 723-734, 2004.

[23] J. Llanes-Coromina, P. Prieto, and P. L. Rohrer, "Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task," in *Proc. 9th Int. Conf. Speech Prosody 2018*, pp. 498-502.

[24] Z. Pei, Y. Wu, X. Xiang, and H. Qian, "The Effects of Music Aptitude and Music Training on Phonological Production in Foreign Languages," *English Language Teaching*, vol. 9, no. 6, p. 19, May 2016. [Online]. Available: https://doi.org/10.5539/elt.v9n6p19

[25] R. V. Teschner and M. S. Whitley, *Pronouncing English: A Stress-Based Approach, with CD-ROM*. Washington, DC: Georgetown University Press, 2004.

[26] S. Weinberger, "Speech Accent Archive," George Mason University, 2015. [Online]. Available: http://accent.gmu.edu.

[27] R. Fuchs, "Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English," in *Proc. 7th Speech Prosody 2014*, Dublin, 2014.

[28] P. Boersma and D. Weenink, "Praat: doing phonetics by computer. Version 6.4.13," 2024. [Online]. Available: http://www.fon.hum.uva.nl/praat/.

[29] Y. Xu, "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis," in *Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013, pp. 7-10.

[30] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2024.

[31] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *J. Stat. Softw.*, vol. 82, no. 13, pp. 1-26, 2017, doi: 10.18637/jss.v082.i13.

[32] R. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, R package version 1.10.2, 2024. [Online]. Available: https://CRAN.R-project.org/package=emmeans.

[33] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.

[34] I. Zeromskaite, "The potential role of music in second language learning: A review article," *J. Eur. Psychol. Stud.*, vol. 5, no. 3, pp. 78-88, 2014.

[35] L. He, *Interlanguage Rhythm*, M.A. thesis, University of Edinburgh, Edinburgh, 2010.

[36] J. Yu, Y. Liao, and Y. Lin, "Study on the rhythm characteristics of Chinese English learners under different spoken task types," *Foreign Languages and Their Teaching*, no. 02, pp. 79-90+147-148, 2022, doi: 10.13458/j.cnki.flatt.004848. [in Chinese]

[37] C. Y. Lee and T. H. Hung, "Identification of Mandarin tones by English-speaking musicians and nonmusicians," *J. Acoust. Soc. Am.*, vol. 124, no. 5, pp. 3235-3248, 2008.

[38] F. Delogu, G. Lampis, and M. Olivetti Belardinelli, "Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers?" *Cognitive Processing*, vol. 7, pp. 203-207, 2006.

[39] K. E. Smayda, B. Chandrasekaran, and W. T. Maddox, "Enhanced cognitive and perceptual processing: a computational basis for the musician advantage in speech learning," *Front. Psychol.*, vol. 6, art. 682, 2015.

[40] M. Besson, J. Chobert, and C. Marie, "Transfer of training between music and speech: common processing, attention, and memory," *Front. Psychol.*, vol. 2, art. 94, 2011.

[41] Peretz, Isabelle, et al. "A Novel Tool for Evaluating Children's Musical Abilities across Age and Culture." Frontiers in Systems Neuroscience, vol. 7, 2013, p. 30.

[42] NAODAO, "Online Hearing Test," [Online]. Available: https://www.naodao.com/ [Accessed: Dec. 15, 2024].