Automated Assessment of Sentence Stress Placement in L2 Chinese Using Deep Learning: An Empirical and Pedagogical Study

Wei Wang[1,3*], Hongmei Wang[2,3], Li Jiang[1]

[1] Zhejiang Normal University

[2] Beijing Normal University

[3] Beijing Affiliated Elementary School of Chaoyang Normal School

Received: June 1, 2026

Revised: June 2, 2026

Accepted: June 4, 2026

Published online: June 10, 2025

To appear in: *International Journal of Advanced AI Applications,* Vol. 2, No. 7 (July 2026)

* Corresponding Author: Wei Wang(wangweiblcu@163.com)

Abstract. This study investigated English-speaking learners' acquisition of Chinese sentence stress in ambiguous contexts using a deep learning-based automated assessment approach. A fine-tuned wav2vec 2.0 model achieved 91.2% agreement with human raters (Cohen's $\kappa = 0.85$). Experiment 1 compared learners' and native speakers' stress identification across semantic contexts. Automated acoustic analysis revealed that learners' accuracy was significantly lower in secondary than primary meaning contexts ($p < 0.1$), while native speakers showed no difference. Experiment 2 compared stress location instruction with general training methods. Automated assessment found no significant difference between the two methods ($p > 0.5$). These findings indicate that learners lack proficiency in semantic-driven stress placement, particularly in secondary meanings, and that instruction focusing solely on stress location is insufficient. Methodologically, this study demonstrates the feasibility of deep learning-based automated prosody assessment. Pedagogically, it recommends integrating prosodic features (pitch, duration) into Chinese language teaching and supports AI-driven Computer-Assisted Pronunciation Training (CAPT) systems.

Keywords: *Deep Learning; Automated Assessment; Sentence Stress; Second Language Acquisition; Chinese Prosody*

1. Introduction

Sentence stress plays a vital role in conveying focal information and highlighting semantic priorities in Chinese. Its precise placement directly determines whether speakers' intentions are accurately understood. However, learners of Chinese as a second language frequently exhibit errors in sentence stress placement. For example, when expressing "I will go to Beijing tomorrow," incorrectly placing the stress on "tomorrow" rather than "Beijing" shifts the emphasis from the destination to the time, leading to semantic misinterpretation. Even learners with foundational Chinese proficiency often struggle with improper stress placement and insufficient pitch variation. Mastering proper stress patterns has thus become a pressing challenge in Chinese phonetics instruction.

Traditional approaches to diagnosing sentence stress errors rely heavily on subjective perceptual judgments by human raters, which are time-consuming, labor-intensive, and prone to inconsistencies. This methodological bottleneck limits both the scalability of instructional feedback and the precision of empirical research.

1.1. Sentence Stress in Chinese as a Foreign Language

For decades, sentence stress teaching has not received sufficient attention in Chinese as a Foreign Language (CFL) instruction. Stress and intonation are the weakest aspects in current teaching practices, which can be attributed to inadequate training and inappropriate teaching methods, with native language interference playing a secondary role [12]. Learners exhibit significant errors in determining stress positions; for instance, in “I will introduce you to it,” students often incorrectly stress one moment [11]. Additionally, intonation teaching remains the most challenging aspect of CFL instruction [15].

From a linguistic perspective, Chinese sentence stress is highly varied, with the same sentence yielding different meanings depending on stress placement—sometimes resulting in completely distinct interpretations. For example, when the phrase “What do you think now?” is stressed on “how,” it asks about temporal changes in thinking style; however, if the stress is placed on “think,” it implies a contrast with prior lack of thinking. Through his teaching practice, Yang observed that international students often struggle to identify appropriate stress positions, ultimately reading sentences with different stress patterns as if they were identical. However, it is important to note that Yang's work primarily focuses on intrinsic aspects of Chinese and does not elaborate on the acquisition process [19].

1.2. Empirical and Pedagogical Studies

In empirical research, international students worldwide were tested and found to exhibit no clear pattern in their stress errors at the grammatical level, advocating for enhanced instruction on stress positioning [21]. Experimental phonetics was employed to examine core stress representation in Chinese declarative sentences among Japanese students, revealing that they frequently use medium-to-low pitch differences to indicate core stress, though this placement often fails to align with focus patterns [14]. Analyzing 48 sentences with varying focal points, it was observed that American students exhibit habitual rhythmic expressions of core stress in specific syllable combinations; yet, core stress frequently does not correspond to actual focus positions [20]. However, this study primarily focused on pitch features, neglecting duration, and did not incorporate native speaker comparative data.

Regarding textbook organization, current CFL textbooks typically incorporate sentence stress instruction only in the beginner-level phonetics section, after which this content is no longer covered. However, modern Chinese features rich grammatical structures related to sentence stress, such as the “shi...de” construction, “lian...dou” (even/also), and “hai...ne” (still/yet). Without teacher guidance, students may grasp basic syntactic structures but struggle to identify key information points; therefore, Cheng advocates extending sentence stress instruction into grammar teaching at intermediate and advanced levels [5].

1.3. Research Gaps and the Present Study

Existing research has elucidated the challenges learners face in producing sentence stress from various perspectives. However, most studies either remain at descriptive observation and empirical summary levels, or focus solely on output characteristics within specific native language contexts, lacking systematic examination of learners' ability to determine stress positions based on semantic information. Furthermore, from a methodological standpoint, nearly all previous studies have relied exclusively on human perceptual judgment for stress evaluation, which suffers from inherent limitations in consistency, scalability, and interpretability.

The potential of deep learning-based automated assessment to address these limitations—by providing consistent, replicable, and fine-grained acoustic analysis—has not yet been explored in L2 Chinese sentence stress research. The present study fills this gap by integrating a fine-tuned wav2vec 2.0 model into the experimental paradigm, enabling consistent evaluation of stress placement accuracy and acoustic feature-level analysis to complement traditional perceptual judgment.

1.4. Research Questions

This study addresses two core questions:

- How do Chinese learners determine sentence stress positions according to different semantic contexts? Specifically, what differences exist between learners' and native speakers' ability to judge stress placement in primary versus secondary semantic contexts?

- What strategies can effectively enhance learners' acquisition of sentence stress? For learners with inadequate judgment skills, which teaching methods prove most effective, and can instruction solely focused on stress placement achieve desired educational outcomes?

2. Methodology

2.1. General Design and Task Procedure

To address the two research questions, this study designed two experiments using ambiguous Mandarin Chinese sentences in which different stress placements yielded different interpretations. Experiment 1 compared English-speaking learners' and native speakers' ability to determine and produce appropriate sentence-stress positions from semantic contexts. Experiment 2 evaluated the effectiveness of a sentence stress-position reinforcement training method through classroom-based practice.

In both experiments, participants produced target sentences according to contextual cues presented on a computer screen. The tasks were programmed and administered using E-Prime 2.0, and oral responses were recorded in mono WAV format at 16 kHz and 16-bit resolution. Stress accuracy was evaluated using both human perceptual judgment and automated classification, as described in Section 2.4.

2.2. Experiment 1

2.2.1. Participants

Experiment 1 included 24 participants. The learner group consisted of 12 advanced English-speaking learners of Chinese, including 7 males and 5 females, aged 20–35 years. They had studied Chinese for approximately five years. The native-speaker control group consisted of 12 native Mandarin speakers, including 6 males and 6 females, aged 22–24 years. Before the experiment, all participants completed a brief oral production task to confirm that they were able to produce continuous Mandarin speech.

2.2.2. Design and Materials

Experiment 1 adopted a 2 × 2 mixed design. The within-subjects factor was context type, with two levels: primary-meaning context and secondary-meaning context. The between-subjects factor was group, with two levels: English-speaking learners and native Mandarin speakers. The dependent variable was sentence-stress placement accuracy.

The experimental materials consisted of 20 ambiguous Mandarin sentences and 4 practice items. Each experimental sentence was paired with two semantic contexts, each requiring a different sentence-stress position and yielding a different interpretation. Thus, each participant completed 40 critical trials. The contexts were classified as primary- or secondary-meaning contexts based on judgments from native Mandarin-speaking raters.

2.2.3. Procedure and Data Analysis

On each trial, participants saw a semantic context and a target ambiguous sentence on the screen. They were asked to produce the sentence naturally while using stress placement to convey the intended meaning. Four practice trials were administered before the formal task.

Sentence-stress accuracy was calculated for each participant under each context condition. A mixed-design ANOVA was conducted with context type and group as independent variables and stress-placement accuracy as the dependent variable. Simple-effects analyses were conducted where appropriate.

2.3. Experiment 2

2.3.1. Participants

Experiment 2 included 20 beginner English-speaking learners of Chinese. They were assigned to either a sentence stress-position reinforcement training group or a comparison group, with 10 participants in each group. Participants were aged 18–35 years and had studied Chinese for approximately one to one and a half years. All participants had basic Chinese grammar and conversational ability and completed a brief oral production task before the experiment.

2.3.2. Design and Materials

Experiment 2 adopted a 2 × 2 mixed design. The between-subjects factor was training method, with two levels: sentence stress-position reinforcement training and general instruction. The within-subjects factor was test type, with two levels: pre-test and post-test. The dependent variable was sentence-stress placement accuracy.

The pre-test and post-test used the same 20 ambiguous sentences as Experiment 1, presented in different orders. Four practice items were included before each test. The training materials consisted of 30 additional ambiguous sentences that did not overlap with the test items.

2.3.3. Procedure and Data Analysis

Participants first completed the pre-test using the same sentence-production task as in Experiment 1. After the pre-test, they completed a two-week instructional period consisting of four training sessions. The reinforcement-training group received explicit instruction and practice on how different stress placements change the interpretation of ambiguous Mandarin sentences. Training activities included instructor demonstration, guided repetition, contrastive practice, rapid-response drills, and immediate feedback. The comparison group received general instruction without specialized sentence-stress reinforcement. The post-test was administered within one to two days after the training period.

Sentence-stress accuracy was calculated separately for the pre-test and post-test. A mixed-design ANOVA was conducted with test type and training method as independent variables and stress-placement accuracy as the dependent variable. The interaction between test type and training method was used to determine whether the reinforcement training produced greater improvement than the comparison condition.

2.4. Stress Accuracy Evaluation and Automated Classification

Stress accuracy was assessed using both human perceptual judgment and automated classification. For human judgment, five native Mandarin Chinese speakers who did not participate in the experiments independently evaluated whether each response carried stress in the position required by the intended interpretation. A response was coded as correct when at least four of the five judges agreed. Inter-rater reliability was calculated using Fleiss' kappa.

Table 1. Corpus construction, annotation, and data split.

Dimension	Description
Corpus type	Self-constructed Mandarin read-speech corpus
Corpus size	5,000 utterances
Corpus composition	250 ambiguous sentence frames × 2 semantic contexts × 10 native Mandarin speakers
Annotation	Binary stress-position annotation by five trained native Mandarin-speaking annotators
Reliability	Fleiss' $\kappa = 0.81$
Data split	Speaker-level split: 3,500 training / 500 validation / 1,000 test utterances

Note. The corpus consisted of read speech rather than natural or synthetic speech. Final labels were determined by majority vote, with unresolved disagreements adjudicated by an expert phonetician.

For automated assessment, a wav2vec 2.0 Base model was fine-tuned to classify the target sentence-stress position. In this study, stress-position classification was defined as a binary task: for each ambiguous sentence frame, the model classified whether the primary sentence stress was realized on the first or the second predetermined candidate position. The fine-tuning corpus was a self-constructed Mandarin read-speech corpus containing 5,000 ambiguous-sentence utterances. Sentence design was informed by previous work on Mandarin stress-related ambiguity and prosodic focus, while the annotation protocol was guided by Mandarin prosodic annotation conventions, particularly the C-ToBI framework [4,10,18]. Details of the corpus construction, annotation procedure, inter-annotator reliability, and data split are summarized in Table 1.

The wav2vec 2.0 Base encoder was followed by a linear classification layer for binary stress-position classification. The main fine-tuning configuration is reported in Table 2. Model performance was evaluated on the held-out test set by comparing predicted stress-position labels with human-annotated labels. The fine-tuned model achieved 91.2% agreement with human raters, with Cohen's kappa = 0.85.

Table 2. Fine-tuning hyperparameters of the wav2vec 2.0 Base model.

Parameter	**Setting**
Model architecture	wav2vec 2.0 Base
Task	Binary classification of target sentence-stress position
Learning rate	2×10^{-5}
Batch size	16
Epochs	20
Optimizer	AdamW
Loss function	Cross-entropy loss

After fine-tuning, the classifier was applied to the experimental recordings. A response was coded as correct by the automated method when the predicted stress position matched the stress position associated with the intended interpretation. For the final analysis, only responses judged correct by both human raters and the automated classifier were treated as accurate. Discrepancies between the two methods were examined separately and are reported in the Results section.

3. Results

3.1. Experiment 1

A two-way ANOVA was conducted with language background and semantic context as independent variables and sentence-stress placement judgment accuracy as the dependent variable. The results showed a significant main effect of language background, $F\,(1,22) = 139.37, p < 0.01$, indicating that native Mandarin speakers outperformed advanced English-speaking learners of Chinese. The main effect of semantic context was also significant, $F(1,22) = 52.19, p < 0.01$, with higher accuracy in primary-meaning contexts than in secondary-meaning contexts. The interaction between language background and semantic context was significant, $F\,(1,22) = 19.34, p < 0.01$.

Simple-effects analyses revealed that native speakers performed significantly better than advanced learners in both primary-meaning and secondary-meaning contexts, with both p-values less than 0.01. Advanced learners demonstrated significantly higher accuracy in primary-meaning contexts compared to secondary-meaning contexts ($p < 0.01$), while native speakers exhibited no significant difference between the two contexts ($p < 0.58$). These results suggest that advanced learners' sentence-stress judgments were more influenced by semantic context, whereas native speakers maintained consistently high accuracy throughout. Refer to Figure 1 for a visual representation of these findings.

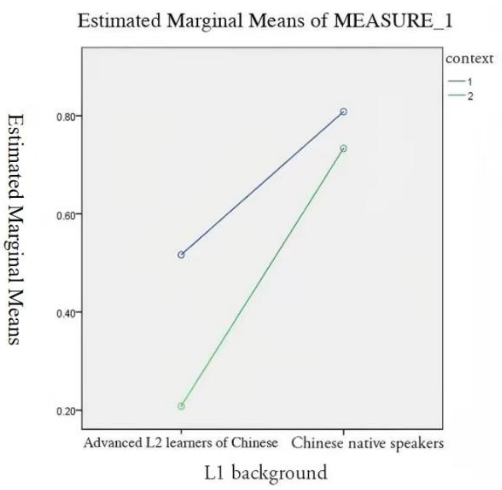

Figure 1. Accuracy rates of sentence stress placement judgment among learners with varying Chinese proficiency levels.

The fine-tuned wav2vec 2.0 model showed 91.2% agreement with human raters, corresponding to an 8.8% disagreement rate, with Cohen's $\kappa = 8.5$. The model-based results were consistent with the human-rating results: advanced learners showed higher accuracy in primary-meaning contexts than in secondary-meaning contexts, $p < 0.01$, whereas native speakers showed no significant context effect, $p < 0.62$. Disagreements mainly involved low-intensity utterances or utterances with non-modal voice quality and were examined separately.

3.2. Experiment 2

A mixed ANOVA was conducted with test type, pre-test vs. post-test, and training method, stress-position reinforcement vs. general instruction, as independent variables and sentence-stress placement accuracy as the dependent variable. The stress-position reinforcement group improved from 62.3% $SD = 8.45$ in the pre-test to 66.8% $SD = 9.12$ in the post-test, while the general instruction group improved from 61.8% $SD = 7.98$ to 64.5% $SD = 8.76$.

The main effect of test type was not significant, $F\ (1,18)\ =\ 1.67, p\ =\ 2.13$. The main effect of training method was also not significant, $F\ (1,18)\ =\ 0.04\ , p\ =\ 8.35$, nor was the interaction between test type and training method, $F\ (1,18)\ =\ 0.10, p\ =\ 7.51$. These results indicate that stress-position reinforcement instruction did not produce significantly greater improvement than general instruction.

The wav2vec 2.0-based assessment showed 90.5% agreement with human raters, corresponding to a 9.5% disagreement rate, with Cohen's $\kappa\ =\ 8.3$. The model-based analysis produced the same pattern of results, with no significant main effect of test type, $F\ (1,18)\ =\ 1.54, p\ =\ 2.30$, no significant main effect of training method, $F\ (1,18)\ =\ 0.04, p\ =\ 8.47$, and no significant interaction, $F\ (1,18)\ =\ 0.10, p\ =\ 7.59$. Thus, both human and automated assessments suggest that instruction focused only on stress-position reinforcement was insufficient to significantly improve beginner learners' sentence-stress production accuracy.

4. Discussion

Sentence stress plays an important role in spoken communication because it helps speakers highlight intended meanings and reduce ambiguity. In Mandarin Chinese, different stress placements may lead to different interpretations of the same sentence. For learners of Chinese as a second language, inappropriate sentence stress may therefore result in utterances that are grammatically correct but pragmatically unclear or semantically misleading. The present study examined English-speaking learners' ability to identify and produce Mandarin sentence stress and further explored the feasibility of using a wav2vec 2.0-based model for automated assessment.

4.1. Semantic Context Effects on Sentence Stress Judgment

Experiment 1 showed that semantic context affected advanced English-speaking learners and native speakers differently. Native Mandarin speakers showed stable performance across primary and secondary semantic contexts, suggesting that they were able to use sentence stress flexibly to disambiguate meaning. In contrast, advanced English-speaking learners performed better in primary contexts than in secondary contexts, indicating that they still had difficulty using sentence stress to identify less frequent or less salient interpretations of ambiguous sentences.

One possible explanation concerns input frequency and usage experience. Usage-based accounts of second language acquisition emphasize that linguistic knowledge is shaped by repeated exposure to form–meaning mappings [6,13]. Primary semantic contexts are likely to be more frequent and familiar in learners' input, whereas secondary contexts are less commonly encountered. As a result, learners may rely more heavily on dominant interpretations and have greater difficulty using prosodic cues to access alternative meanings. Native speakers, by contrast, have accumulated richer experience with context-dependent stress patterns and can therefore identify the intended meaning more accurately even in secondary contexts.

These findings suggest that advanced learners' difficulty does not lie only in perceiving acoustic prominence, but also in mapping stress placement onto contextually appropriate meanings. Sentence stress instruction should therefore not be limited to prosodic form alone; it should also strengthen learners' awareness of the semantic and pragmatic functions of stress.

4.2. Deep Learning-Based Automated Assessment

A methodological contribution of this study is the use of a wav2vec 2.0-based model to assess L2 Mandarin sentence stress. The model showed high agreement with human raters, reaching 91.2% agreement in Experiment 1, with Cohen's $\kappa = 8.5$, and 90.5% agreement in Experiment 2, with Cohen's $\kappa = 8.3$. These results suggest that the automated assessment was broadly consistent with human judgment and can provide a reliable supplementary tool for evaluating sentence-stress placement.

Compared with traditional perceptual judgment alone, the automated model offers several advantages. It provides consistent scoring across large numbers of responses and reduces potential rater fatigue or rater drift. It also makes it possible to examine systematic disagreement between human and machine judgments. In the present study, disagreement cases were mainly associated with low-intensity utterances or non-modal voice quality, where human raters tended to be more tolerant than the model. These cases indicate that current speech models remain sensitive to atypical voice quality and weak acoustic cues, which should be considered in future model improvement.

More broadly, the findings demonstrate the feasibility of applying deep learning-based assessment to L2 prosodic features. While automated pronunciation assessment has often focused on segmental accuracy or lexical tone, the present results suggest that sentence-level prosodic features such as stress placement can also be evaluated automatically with reasonable reliability.

4.3. Pedagogical Intervention for Sentence Stress

Experiment 2 examined whether explicit reinforcement of sentence-stress position could improve beginner learners' production of Mandarin sentence stress. The results showed no significant main effect of test type, no significant main effect of training method, and no significant interaction between test type and training method. This indicates that the sentence stress-position reinforcement method did not lead to significant improvement from pre-test to post-test.

Several factors may explain this null effect.

First, stress-based ambiguity may have been too difficult for beginner learners within a short instructional period. Learners had to process sentence meaning, identify the intended focus, and produce appropriate prosodic cues at the same time, which may have exceeded their processing capacity.

Second, the training may have focused too narrowly on stress location without sufficient practice of the acoustic cues that realize stress, such as pitch movement, duration, and intensity.

Third, the total training time was limited, with only four short sessions, which may not have been enough to produce measurable gains.

These results suggest that sentence stress instruction for beginner learners should be more gradual and integrated. Rather than only asking learners to identify where stress should fall, future instruction should combine stress-position awareness with acoustic training, including pitch contour, duration lengthening, and intensity control. Longer and more intensive training may also be necessary before clear production gains can be observed.

4.4. Implications for AI-Driven CAPT Systems

The findings have implications for the development of AI-driven Computer-Assisted Pronunciation Training systems. Most current CAPT tools for Mandarin focus mainly on segmental pronunciation or lexical tone, while sentence-level prosodic features receive less attention. The wav2vec 2.0-based model used in this study could be incorporated into CAPT systems to provide automatic feedback on sentence-stress placement.

Such advanced systems could effectively detect whether learners appropriately place stress on the intended sentence positions. They can also display visual feedback on pitch and duration patterns, while intelligently adjusting task difficulty according to individual learner performance. For instance, training could commence with primary semantic contexts, gradually introducing secondary contexts as learners become more familiar with the intricate relationship between stress placement and sentence meaning. This type of adaptive feedback may significantly enhance learners' ability to connect prosodic form with communicative function in a more effective and meaningful way. Ultimately, this approach aims to foster deeper understanding and improve overall communication skills.

4.5. Limitations and Future Directions

Several limitations should be noted. First, the sample sizes in both experiments were relatively small, which may limit the generalizability of the findings. Second, the wav2vec 2.0 model was fine-tuned on read speech produced by a limited number of speakers; its performance on spontaneous speech, noisier recordings, or more diverse accents remains to be tested. Third, the intervention in Experiment 2 was brief, and longer-term training may be needed to determine whether sentence-stress instruction can produce stable improvement. Fourth, the study focused only on English-speaking learners of Chinese, so future research should examine whether similar patterns appear among learners from other first-language backgrounds.

Future studies could address these limitations by recruiting larger and more diverse samples, expanding the speech corpus used for model training, and testing the automated assessment model in classroom-based CAPT systems. Longitudinal intervention studies are also needed to examine whether integrated training of stress position, pitch, duration, and intensity can more effectively improve learners' sentence-stress production.

5. Conclusion

This study investigated English-speaking learners' acquisition of Mandarin sentence stress from two perspectives: their ability to identify sentence-stress positions in different semantic contexts and the effectiveness of strengthened instruction on sentence-stress positioning. The results of Experiment 1 showed that advanced learners were able to identify sentence-stress positions relatively accurately in primary-meaning contexts, but their performance declined in secondary-meaning contexts. In contrast, native Mandarin speakers showed consistently high accuracy across both context types. This finding suggests that semantic context plays a more decisive role in learners' stress judgment than in native speakers' judgment, and that secondary-meaning contexts remain particularly challenging even for advanced learners.

Experiment 2 further showed that instruction focusing only on stress-position reinforcement did not significantly improve beginner learners' sentence-stress production accuracy. This result indicates that sentence-stress acquisition cannot be achieved merely by teaching learners where stress should be placed. Accurate production also requires control of relevant prosodic features, including pitch, duration, and intensity. Therefore, sentence-stress instruction should combine explicit explanation of stress placement with auditory discrimination, imitation, acoustic feature training, and communicative practice.

Methodologically, this study integrated a wav2vec 2.0-based automated assessment model into the evaluation of Mandarin sentence stress. The model showed high agreement with human raters, reaching 91.2% in Experiment 1 and 90.5% in Experiment 2. This suggests that deep learning-based assessment can provide a reliable supplementary tool for evaluating L2 Mandarin sentence-stress production. Disagreement cases were mainly associated with low-intensity utterances or non-modal voice quality, indicating directions for future improvement of the model.

Pedagogically, the findings suggest that sentence-stress instruction should not be limited to beginner-level phonetic training. Instead, it should be incorporated throughout the Chinese language curriculum, especially in intermediate and advanced courses involving grammar, discourse, and oral communication. Teaching materials should also include more activities that connect stress placement with semantic focus and communicative intention.

The study also has implications for AI-driven Computer-Assisted Pronunciation Training systems. The validated automated assessment model could be used to provide real-time feedback on learners' sentence-stress placement and to support adaptive training tasks. Future research should expand the participant sample, include more diverse speech data, improve model robustness, and examine the effectiveness of AI-assisted sentence-stress training in classroom-based longitudinal studies.

In conclusion, English-speaking learners of Chinese still face difficulties in using semantic context to determine Mandarin sentence stress, particularly in secondary-meaning contexts. Instruction that focuses only on stress location is insufficient; more comprehensive prosodic training is needed. At the same time, deep learning-based automated assessment offers a promising approach for evaluating and supporting the acquisition of Mandarin sentence stress.

Acknowledgements

This work was supported by the 2023 Special Project on the Cultivation of Top Innovative Talents under the "14th Five-Year Plan" for Educational Science in Chaoyang District, Beijing (Grant No. 2023ZX009).

References

[1] Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing*

Systems, 33, 12449–12460.

[2] Cao Wen. (2010). 汉语焦点重音的韵律实现 [The prosodic realization of focus stress in Chinese]. *Beijing: Beijing Language and Culture University Press*. [in Chinese]

[3] Chen Yongming, & Cui Yao. (1997). 汉语歧义句的加工 [The processing of ambiguous sentences in Chinese]. *Acta Psychologica Sinica*, (1). [in Chinese]

[4] Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36(4), 724–746.

[5] Cheng Shuqiu. (2005). 句重音在对外汉语语法教学中的作用 [The role of sentence stress in teaching Chinese grammar as a foreign language]. *Continuing Education Research*, (4), 126–128. [in Chinese]

[6] Ellis, N. C. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 297–339.

[7] Ellis, R. (1994). The study of second language acquisition. Oxford: Oxford University Press.

[8] Feng Shengli. (2009). 汉语的韵律、词法与句法：修订本 [Prosody, morphology, and syntax in Chinese: Revised edition]. *Beijing: Peking University Press.* [in Chinese]

[9] Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67–83.

[10] Gao, J., & Gu, Y. (2024). Same sentences, different meanings: Prosodic and gestural resolution of ambiguity in Mandarin Chinese. In Proceedings of Speech Prosody 2024 (pp. 886–890). *International Speech Communication Association*.

[11] Ji Xiuqing. (1998). 语句重音与口语教学 [Sentence stress and oral Chinese teaching]. In 对外汉语语音与语音教学研究 [Studies on Chinese phonetics and phonetics teaching as a foreign language] (pp. 357–368). *Beijing: Sinolingua Press*. [in Chinese]

[12] Lu Jianji. (1984). 中介语理论与外国人学习汉语的语音偏误分析 [Interlanguage theory and an analysis of phonetic errors in foreigners' learning of Chinese]. *Language Teaching and Linguistic Studies*, (3). [in Chinese]

[13] MacWhinney, B. (2006). Emergentism: Use often and with care. *Applied Linguistics*, 27(4), 729–740.

[14] Qiu Shanshan. (2007). 日本留学生汉语陈述句核心重音的韵律表现 [The prosodic realization of core stress in Mandarin declarative sentences by Japanese international students]. *Master's thesis, Beijing Language and Culture University*. [in Chinese]

[15] Wang Yunjia. (2002). 汉语语音研究与汉语语音教学接口的问题 [Issues at the interface between Chinese phonetic research and Chinese pronunciation teaching]. In 对外汉语论丛（第二辑） [Essays on teaching Chinese as a foreign language, Vol. 2]. *Shanghai: Shanghai Foreign Language Education Press*. [in Chinese]

[16] Wang Yunjia, Chu Min, & He Lin. (2006). 汉语焦点重音和语义重音分布的初步试验研究 [A preliminary experimental study of the distribution of focus stress and semantic stress in Chinese]. *Chinese Teaching in the World*, (2). [in Chinese]

[17] Witt, S. M., & Young, S. J. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, 30(2–3), 95–108.

[18] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27(1), 55–105.

[19] Yang Defeng. (1991). 重音与语义：句子重音 [Stress and semantics: Sentence stress]. In 第三届国际汉语教学讨论会论文选 [Selected papers from the Third International Conference on Chinese Language Teaching]. *Beijing: Peking University Press*. [in Chinese]

[20] Zhang Juan. (2009). 美国留学生汉语陈述句核心重音的韵律表现研究 [A study on the prosodic realization of core stress in Mandarin declarative sentences by American international students]. *Master's thesis, Beijing Language and Culture University*. [in Chinese]

[21] Zhu Chuan. (Ed.). (1997). 汉语语音学习对策 [Strategies for learning Chinese phonetics]. *Beijing: Yuwen Press*. [in Chinese]