THE PROSODY-GESTURE COORDINATION IN CHINESE EFL LEARNER' ACADEMIC ORAL PRESENTATION

Yating Cao[1,2] Hua Chen[1]

[1]Nanjing University [2]University of Warwick
yatingcao@smail.nju.edu.cn; chenhua@nju.edu.cn

ABSTRACT

Human communication is multimodal. However, few studies have examined how L2 learners, particularly those whose L1 is a tonal language, coordinate their beat gestures and prosody in an intonation-language speech production. This study explored how Chinese EFL speakers coordinated pitch accents and beat gestures in academic oral presentation and the effects of speakers' language proficiency. Results revealed that the intermediate-level Chinese EFL learners produced beat gestures regardless of the placement of pitch accents and the information status of their accompanying speech. Beat gestures were used to fulfill the meta-linguistic function. Whereas, proficient learners produced significantly fewer beat gestures, and beat gestures were more likely to co-occur with pitch accents to stress the crucial information of discourse for its meta-pragmatic function. These findings provide empirical evidence from L2 learners to expand the literature on multimodal prosody and will have pedagogical implications for English teaching.

Keywords: beat gestures, pitch accent, gesture-speech coordination, proficiency level

1. INTRODUCTION

Speech communication is multimodal in nature. As one of the crucial non-verbal cues, gestures often accompany people's speech production, providing another window into people's minds [1-3]. McNeill [3] postulated three synchrony rules for speech and gestures, namely semantic, pragmatic and phonological synchrony rules. The first two refer to that gestures often parallel meaning and pragmatic functions with speech, and the last one, phonological synchrony, which is also the most pertinent to the present study, states that the stroke phase of gestures tends to occur at or just before the stressed syllable of the lexical affiliate. Here, the stroke means the only obligatory and the most prominent part of a gesture [2-3].

Pitch accent is one type of prosodic specification where a segment (a syllable or a word) within a sentence stands out from its context to convey focus or information status [e.g., 4-5]. In more recent years, a growing body of research has consistently reported a close coordination between prosodic prominence and manual gestures [e.g., 7-10]. Beat gestures, according to [3], refer to "a simple flick of the hand or fingers up and down; or back and forth". This type of gesture has been found to be the most frequently used gesture in daily conversation, particularly taking up 94.6% of gesture types in academic-style conversations [11]. And it has been believed that beat gestures, like pitch accent, is usually used to emphasize the information status as a "gestural yellow highlight" [12].

There is evidence that beat gestures and pitch accent are closely coordinated. The preliminary analysis by [13] investigated how different gesture types were synchronized with English speech, and showed a strong coupling between beat gestures and pitch accent. For example, 86% of beat gestures produced by one Australian English speaker co-occurred with a pitch accent. Shattuck-Hafnagel & Ren [11] analyzed 1,334 beat gestures in a 20-min academic speech and found a close coordination between beat gestures and the prosodic structure of spoken utterances.

While the phonological coordination between prosodic and gestural prominence is well established in prior research, it is mainly based on studies of non-tonal languages, such as English [14], Catalan [7], Dutch [15], and French [16]. It would be interesting to ask whether and how prosody and gestures coordinate in non-stress tone languages. Fung et al. [17] examined how ten native speakers of Hong Kong Cantonese coordinated prosody and deictic gestures in a picture-verification task. They not only found a close coordination between prosodic and gestural prominence but also revealed that the prosodic anchor was not F0 as previous studies showed but durational increase instead, which manifested cross-linguistic similarities and differences.

Furthermore, it should also be noted that most research has been focusing on how native speakers achieve prosody-gesture coordination, while research on the patterns of non-native speakers has been quite sparse. McCafferty [18] examined EFL learners' use of beat gestures in conversation and suggested that these gestures were the materialization of their inner prosody. Further, he suggested that L2 learners employed beat gestures for a meta-linguistic purpose, and that native speakers were more likely to use beat

gestures to fulfill its meta-pragmatic function. This research strengthens McNeill's [3] argument by extending it beyond native speakers to L2 learners. Graziano et al. [19] explored how English and French monolinguals and bilinguals achieved phonological coordination in a cartoon retelling task. Results showed that there was no significant difference concerning the time-aligned coordination of gestures and speech, which is possibly explained by the relatively high level of proficiency in the two languages. In light of the findings, it would be an important aspect to further test participants at different proficiency levels to better inform the nature of gesture-speech links and theories of language development as well. Interestingly, as mentioned before, Mandarin Chinese is a tonal language that exhibits the syllable-timed rhythmic pattern; hence, Chinese EFL learners might accentuate every word within a sentence without clear communicative intentions as a result of the L1 transfer [20]. In this sense, how this group of learners coordinates prosodic and gestural prominence in their L2 warrants further investigation.

To respond to these gaps, the present study took the preliminary attempt to analyze the academic oral English presentation by Chinese EFL learners and addressed the following questions: (1) How do beat gestures align with prosodic prominence in Chinese EFL learners' speech production? (2) Is there any difference between the two proficiency groups?

2. METHODS

2.1. Data source

Multimodal Corpus of Academic English Speech by Chinese Learners (MCAESCL) [21] was served as data source in this study. All the participants in this corpus were native Chinese speakers, and they were required to provide a two-minute English academic oral presentation. There were no other audiences other than the researcher in the experiment setting, and the whole process was captured by a high-definition camera. Also, they were randomly assigned to topics, that is "*How to prepare for an oral presentation*?" for the informative speech genre or "*To read selectively or extensively? What is your understanding? Support your ideas with details*" for persuasive speech. Both topics were relevant to students' college academic life so that they could express their opinions [see 21 for further review].

So far, MCAESCL consists of 1,763 speech samples, with varying majors, academic levels of speakers' university, and gender balanced. Among all these data, 924 speakers' speech productions were accompanied by gestures and became the candidate data for the studies related to gestures. Besides, all the speech samples have been rated by 15 linguistic experts from various linguistic dimensions (i.e., phonetics, syntax, vocabulary, and discourse) via focused analytical scoring, and all the results have been validated through multiple methods. Based on the total score, all the samples could be ranked and categorized into different proficiency levels (i.e., advanced, intermediate, and low).

Due to the exploratory nature of the present study the sample size is quite modest. 12 speech samples were selected, with half from the advanced learner (AL) and intermediate learner (IL) group, respectively. In addition, all the chosen speakers gave speeches of the same topic and achieved gender balance.

2.2. Annotation

The data annotation consisted of two parts – prosody and gestures. Previous studies have shown that the two modalities, namely audio and visual mode, might influence the prominence perception [e.g., 11, 22]; therefore, these two parts of annotations were independently completed to eliminate possible effects.

For the prosodic annotation, the speech was initially transcribed orthographically and labeled for its intonational structure – mainly for pitch accents, using Praat [23] and ToBI as the prosodic annotation system. And all the gesture annotation was accomplished using frame-by-frame analysis in ELAN [24]. There were three main tiers used in the current coding scheme as follows:

- Main Tier 1: orthographic transcription
- Main Tier 2: Manual gesture phase
- Main Tier 3: Prosodic information

Specifically, for the Main Tier 2, the present study followed [25]'s proposal for coding the manual gesture phase, which included preparation, beginning, and end of the gesture stroke, retraction phase, as well as optional pre- or post-stroke recovery. In addition, a sub-tier was coded for the gesture apex separately to indicate the gestural prominence. The Main Tier 3 was coded after the completion of the first two tiers, which contains two main sub-tiers. The first sub-tier comprised already annotated ToBI labelling imported from Praat. And the second sub-tier was used to label whether the gestural prominence was linked with prosodic prominence or not, suggesting by the symbol (+) and (-).

Consistent with prior research [26], the annotation was carried out by the first author and then the second coder, who is an experienced coder of gestures, was invited to annotate 25% of randomly-selected data. Then, the Cohen's Kappa value (k=0.82) indicated the high inter-annotator reliability.

2.3. Data analysis

The present study calculated the percentage of co-occurrence of gestural prominence (i.e., gesture apex) in beat gestures and prosodic prominence (i.e., pitch accents). The chi-square analysis was then carried out to examine whether there existed a difference between the two proficiency levels (i.e., AL vs. IL) concerning the percentage of occurrence of gestural prominence and prosodic prominence.

3. RESULTS

Table 1 displays how Chinese EFL learners use beat gestures to achieve the coordination between prosodic and gestural prominence. It could be seen that advanced learners used more gestures than intermediate-level learners did in their academic oral presentations. Notably, beat gestures took a large proportion of the total gesture use as shown by the proportion. In addition, Chinese EFL learners at the intermediate level almost only used beat gestures throughout their speech production, exceding more than 83% of the total gesture types. However, learners of this group seldom coordinated beat gestures with pitch accents in spoken language.

Further chi-square analysis was performed to examine whether there existed differences between speakers of language proficiency levels in terms of the co-occurrence of prosodic and gestural prominence. Results revealed a significance in these two groups (χ^2 = 53.78, df = 1, p = 0.0001), with advanced learners showing more coordination.

Speakers	Gestures (Total)	Beat gestures (%)	The coordination between prosodic prominence and gestural prominence (%)
AL 1	43	22 (51.17%)	10 (45.5%)
AL 2	30	14 (46.67%)	9 (64.3%)
AL 3	47	38 (80.86%)	20 (52.64%)
AL 4	39	27 (69.23%)	17 (62.96%)
AL 5	47	35 (74.47%)	21 (60%)
AL 6	43	21 (48.83%)	15 (71.43%)
IL 1	49	43 (87.76%)	13 (30.23%)
IL 2	19	19 (100%)	3 (15.79%)
IL 3	33	29 (87.88%)	9 (31.03%)
IL 4	44	40 (90.91%)	4 (10%)
IL 5	43	36 (83.72%)	8 (22.22%)
IL 6	24	20 (83.33%)	5 (25%)

Table 1: The gesture use and coordination between prosodic prominence and gestural prominence in Chinese EFL learners' academic oral presentation.

4. DISCUSSION

The present study has two research goals. The first one is to show how Chinese EFL learners coordinate their prosodic prominence and gestural prominence (beat gestures) to construct meaning in academic oral presentation. And the second one is to examine the role of language proficiency in this issue.

Firstly, the current study found that beat gestures constituted a large proportion of Chinese EFL learners' speech production. This finding offered evidence from Chinese EFL learners to [11] that beat gestures are the most frequently used gesture type in native English speakers' academic-style report. In fact, Alibali et al. [27] compared speakers' gesture use (representational vs. beat gestures) when they could see the interlocutors or not. Results showed that the number of beat gestures was not affected as much as representational gestures under the invisibility condition, which indicated that beat gestures were more likely to be used for self-regulatory functions. In the present study, speakers might use beat gestures to accompany their speech to help themselves to regulate the speech flow as their "inner prosody" [18]. Another important factor worth consideration is the task effects. Previous studies usually employed the cartoon-retelling task to elicit the gestures, where certain gesture types such as iconic gestures and metaphoric gestures are used much more often. However, as Tabensky [28] suggested, different speech genres may influence how speakers use gestures to convey meaning. The academic oral presentation used in this study is more abstract compared to other tasks, which might be challenging for speakers to produce more meaning-based gestures.

Second, even though the proportion in the present study was not big enough, it still found that Chinese EFL learners could coordinate gestural prominence and prosodic prominence to construct discourse meaning in their academic oral English presentation. Since previous studies has mainly focused on the non-tonal languages [7, 14-16], the current study provided evidence that as tonal language speakers, Chinese speakers also achieved the prosody-gesture coordination in their L2 speech (i.e., English). This evidence also shows that L2 speakers would consider speech and gestures in their speech production planning process to form an integrated system [1-3], suggesting a universality of embodiment in human communication. However, it should be noted that the present study only examined how Chinese speakers coordinated prosodic and gestural prominence in their English production, while no further investigation of how they performed in their L1 speech (i.e., Mandarin Chinese). Therefore, these speakers' ability to achieve such coordination could simply because they

were speaking English, which is found to be closely linked between speech and gestures in production.

Third, the proficiency level of speakers was found to be an important factor in prosody-gesture coordination [19]. In the present study, advanced Chinese EFL learners were more likely to use the apex of beat gestures in conjunction with pitch accents to emphasize important information, thereby achieving embodied communication. In contrast, intermediate-level Chinese EFL learners tended to use gestures without clear communicative purposes, using beat gestures for every word or even syllable. This group difference is similar to what McNeill [3] observed in the beat gesture production of native English-speaking adults and children. Specifically, due to the higher English proficiency of advanced English learners, they were more likely to perform like native speakers by better controlling their speech planning process, where beat gestures were used for pragmatic purposes. Whereas, intermediate-level Chinese EFL learners may experience higher cognitive load during speech production, resulting in weaker ability to convey their communicative intentions through a combination of prosodic and gestural devices. Thus, beat gestures were used for a meta-linguistic function, similar to what native children do [3]. Another possible factor is the different rhythmic pattern between Chinese and English. Since Chinese is a syllable-timed language, where every syllable is stressed to convey meaning, intermediate-level Chinese EFL learners might be negatively influenced by their native language and tend to stress each word or syllable in English production. Consequently, the usage of beat gestures may simply reflect the "materialization of their syllable-timed rhythmic pattern in spoken English," as McCafferty [18] stated. However, this conclusion should be made with caution since the present study did not directly examine how Chinese EFL learners produce prosodic and gestural prominence in their L1. Further research is needed to explore whether the pattern of intermediate learners is more similar to that of Chinese native speakers and the pattern of advanced learners is more like that of English native speakers.

5. CONCLUSION

The present study investigates how Chinese EFL learners coordinate the prosodic prominence (pitch accents) and gestural prominence (apex of beat gestures) in academic English oral representation. Results showed that L2 learners can also use their gestures as a tool to construct discourse as native speakers did, giving empirical evidence for the integration of speech and gesture in L2 speakers. However, the coordination patterns may vary depending on the level of L2 proficiency of the speaker. Chinese EFL learners at advanced level were more likely to coordinate prosodic prominence and gestural prominence to convey their communicative intentions. The findings of the study could offer L2 evidence to the literature on speech-gesture coordination and provide positive implications for multimodal English teaching and learning.

However, this study was still the preliminary attempt to tap into the phonological coordination of beat gestures and pitch accents for the smaller sample size. In addition, further analysis of phonological coordination patterns also merits further investigation to advance the understanding of this field. For example, it is interesting to explore prosodic anchoring in such phonological coordination in Chinese EFL learners' speech. Additionally, further investigation is needed to understand how Chinese EFL learners coordinate prosodic and gestural prominence in their native language, in order to better explain the observed group differences. The final note is that the task used in the present study was academic oral presentation, which differs from the commonly-used cartoon-retelling task. Therefore, future research may consider the task effects on L2 learners to further understand the prosody-gesture coordination.

6. ACKOWNLEDGEMENTS

This study was supported by the National Social Science Foundation of China (20AYY013), and China Scholarship Council (202206190122).

7. REFERENCES

[1] Goldin-Meadow, S. 2003. *Hearing Gesture: How Our Hands Help us Think*. Harvard University Press.

[2] Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.

[3] McNeill, D. 1992. *Hand And mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

[4] Bishop, J., Kuo, G., Kim, B. 2020. Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics* 82, 100977.

[5] Cutler, A. 1984. Stress and accent in language production and understanding. In: Gibbon, D., Richter, H. (Eds.). *Intonation, Accent and Rhythm* Berlin: De Gruyter, 77-90.

[6] Cutler, A. 2005. Lexical stress. In: Pisoni D. B., Remez, R. E. (Eds.). *The Handbook of Speech Perception*. Malden, MA: Blackwell, 264–289.

[7] Esteve-Gibert, N., Pons, F., Bosch, L., Prieto, P. 2014. Are gesture and prosodic prominences always coordinated? Evidence from perception and production. *Proceedings of the International Conference on Speech Prosody 2014*, Dublin, Ireland, 222-226.

[8] Krahmer, E., Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57(3), 396–414.

[9] Krivokapić, J., Tiede, M. K., Tyrone, M. E. 2017. A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory Phonology* 8(1), 3.

[10] Wagner, P., Malisz, Z., Kopp, S. 2014. Gesture and speech in interaction: An overview. *Speech Communication* 57, 209–232.

[11] Shattuck-Hufnagel, S., Ren, A. 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology* 9.

[12] McNeill, D. 2006. Gesture: A psycholinguistic approach. *The Encyclopedia of Language and Linguistics*, 58–66.

[13] Yasinnik, Y., Renwick, M., Shattuck-Hufnagel, S. 2004. The timing of speech-accompanying gestures with respect to prosody. *Journal of the Acoustical Society of America* 115(5), 2397.

[14] Loehr, D. P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3(1), 71–89.

[15] De Ruiter, J. P. 1998. *Gesture and Speech Production*. Unpublished Doctoral Dissertation. Nijmegen: Radboud University.

[16] Ferré, G. 2014. A multimodal approach to markedness in spoken French. *Speech Communication* 57, 268-282.

[17] Fung, H. S. H., Mok, P. P. K. 2018. Temporal coordination between focus prosody and pointing gestures in Cantonese. *Journal of Phonetics* 71, 113–125.

[18] McCafferty, S. G. 2006. Gesture and the materialization of second language prosody. *International Review. of Applied Linguistics in Language Teaching* 44(2), 179–207.

[19] Graziano, M., Nicoladis, E., Marentette, P. 2020. How referential gestures align with speech: Evidence from monolingual and bilingual speakers. *Language Learning* 70(1), 266–304.

[20] Chen, H., Wang, X. 2015. A study of English phrasal accentuation in Chinese EFL learners'read speech. *Foreign Languages and Their Teaching* 03, 13-18+40.

[21] Chen, H., Cao, Y., Ma, D. 2022. MCAESCL-A Multimodal Corpus of Academic English Speech by Chinese Learners. *Technology Enhanced Foreign Language Education* 3, 45-51+109.

[22] Prieto, P., Cravotta, A., Kushch, O., Rohrer, P. L., Vilà-Giménez, I. 2018.Deconstructing beat gestures: A labelling proposal. *Proceedings of the 9th International Conference in Speech Prosody 2018*, 201-205.

[23] Boersma, P., Weenink, D. 2022. Praat: doing phonetics by computer. Version 6.2.06, https://www.praat.org.

[24] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. 2006. ELAN: A professional framework for multimodality research. *In Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy. http://www.lrec-conf.org/proceedings/lrec 2006/.

[25] Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., Demuth, K. 2016. Non-referential gestures in adult and child speech: Are they prosodic? *Speech Prosody 2016*, 836–839.

[26] Parrill, F. 2010. The hands are part of the package: Gesture, common ground, and information packaging. In: Rice, S., Newman, J (Eds.). *Empirical and Experimental Methods in Cognitive/Functional Research.* Stanford Publications, 285-302.

[27] Alibali, M. W., Heath, D. C., Myers, H. J. 2001. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language* 44, 169–188.

[28] Tabensky, A. 2008. Expository discourse in a second language classroom: How learners use gesture. In S. McCafferty & G. Stam (Eds.), *Gesture: Second language acquisition and classroom research*. London, UK: Routledge. 298—320.