Learning L2 Prosody using Gestures: The Role of Individual Differences related to Musicality

Lieke van Maastricht[1], *Marieke Hoetjes*[1], *Lisette van der Heijden*[1]

[1]Centre for Language Studies, Radboud University, Nijmegen
lieke.vanmaastricht@ru.nl, marieke.hoetjes@ru.nl

Abstract

The present study aimed to disentangle the influence of gesture type, physical involvement level, and individual differences in learner characteristics, i.e., working memory (WM) capacity and musicality, in determining the effectiveness of L2 lexical stress training. To this end, 60 native speakers of Dutch read aloud Spanish phrases containing cognates, which were counterbalanced for lexical stress position compared to their Dutch counterpart (e.g., 'piRÁmides' in Spanish, 'piraMIdes' in Dutch). They did so as a pre-test before receiving lexical stress training (T1) and as a post-test both directly after training (T2), and approximately one hour later (T3). Subjects received lexical stress training in one of five conditions varying in gesture type and physical involvement level: audio-visual (AV), AV-beat-perception, AV-beat-production, AV-metaphoric-perception, AV-metaphoric-production. Between T2 and T3, subjects performed a WM capacity and musical aptitude task. The results show that irrespective of training condition subjects significantly improved their L2 lexical stress production from T1 to T2 and T3. Although differences between training conditions were non-significant, there were several significant three-way interactions between WM capacity or musical aptitude and testing time and training condition. This underlines the importance of considering task and learner characteristics in determining the gestural benefit in learning L2 prosody.

Index Terms: multimodality, gesture, L2 acquisition, lexical stress, prosody, individual differences

1. Introduction

Previous studies have demonstrated both the integrated relation between speech and gesture in communication [1-2] and the beneficial role that gestures can play in first language (L1) acquisition [3]. However, theories about the use of gestures in foreign language (L2) learning are less conclusive. Although several studies have demonstrated the benefit of using gesture in L2 vocabulary learning [4-6], little is known about the effectiveness of gestures in other linguistic areas. Since the gestural benefit in L2 vocabulary learning might be explained by the close semantic relation between speech and gesture [7], one may wonder whether gestures also improve L2 learning when the speech-gesture relation is not semantically based, like in L2 prosody learning, which comprises the acquisition of phrasal and lexical stress, intonation, and rhythm [8].

Exploring the effect of gestures in learning L2 prosody is interesting for two main reasons. First, in L2 learning, prosody contributes to L1 perceptions of (non-)nativeness [9-11], and, since achieving native-like pronunciation is still considered the norm [12], L2 speakers who make prosodic errors seem to suffer several negative consequences [13-14]. Hence, if gestures would help in learning L2 prosody this could be a theoretically valid and practical addition to L2 teaching methods. Second, beat gestures, which commonly visualise speech rhythm, are temporally aligned with prosodically prominent elements in speech [15-17]. Thus, in natural speech, there appears to be a direct relation, though not a semantic one, between beat gestures and prosodic prominence. While prior research has reported some positive trends for the effect of gesture on L2 prosody learning [18-19], findings are often barely significant [20-21], and studies reporting no gestural benefit also exist [22-23]. A potential explanation for these varying findings is that factors like gesture type (i.e., beat or metaphoric), physical involvement level (i.e., producing or perceiving gestures), and learner characteristics might be important in determining the effectiveness of gestures in L2 prosody training.

First, different gesture types may affect L2 prosody acquisition in different ways. Beat gestures, for example, have a natural relation with prosodic prominence in speech [15], while metaphoric gestures might better visualise specific L2 prosodic contrasts (e.g., the rising/falling of Mandarin tones [24]). Second, the physical involvement level during training might also affect learning outcomes: Producing gestures involves a more embodied representation and an additional modality compared to perceiving gestures, hence according to theories of embodied cognition and multimodality, it is expected that producing gestures results in a greater learning benefit than only perceiving them [7, 25-26]. Third, individual learner characteristics might affect the relationship between gestures and L2 prosody learning too. [27] reported that the benefit of perceiving and producing gestures during language learning depends on learners' cognitive abilities, such as their WM capacity. Musical aptitude, the ability to hear patterns in sets of sounds, might be another important learner characteristic in this context, as it is closely related to prosodic learning [28]. Although, separately, gesture type, physical involvement level, and individual differences concerning musicality appear to be relevant in determining the gestural benefit in L2 prosody learning, it is still unknown how they interact when combined. Hence, this paper attempts to untangle the effect of gesture type, physical involvement level, WM capacity, and several musical aptitude measures in the context of L2 lexical stress training.

In our experiment, Dutch subjects were trained to produce Spanish lexical stress. Learners generally struggle with lexical stress in their L2, especially in cognates that are highly similar except for their stress position (e.g., 'proFEssor' in Dutch, but 'profeSOR' in Spanish). Our participants were trained either with beat gestures, metaphoric gestures, or without gestures. Moreover, the training conditions varied in physical involvement level as subjects were asked to either produce or

10.21437/SpeechProsody.2022-146

perceive the gestures during training. Additionally, productions were measured at three different time points (i.e., pre-test; T1, immediate post-test; T2, and delayed post-test; T3). Finally, WM capacity and musical aptitude were measured using a backwards digit span task [29] and a music perception task using subtests from the Profile of Music Perception Skills (PROMS) test battery [30].

2. Method

2.1. Participants

Sixty Dutch natives participated in the study (45 females, *M* age = 23.86 years, *SD* = 8.68), almost all were raised monolingually and had no, or little, knowledge of Spanish. They received 10 euros or course credits after participating and their age, gender, language, and education background did not differ across the five experimental conditions. Participants were randomly assigned to one of the five conditions.

2.2. Materials

The study contained 1) a read-aloud task, 2) lexical stress training, 3) a musical aptitude task, 4) a working memory task, and 5) a questionnaire.

2.2.1. Read-aloud task

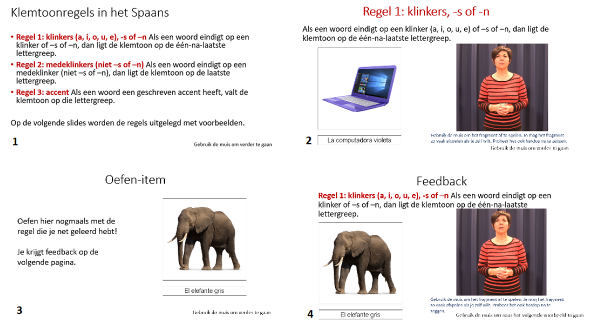
At T1, T2, and T3, participants read aloud 31 short Spanish phrases (e.g., *El elefante gris*, the grey elephant) containing a Spanish-Dutch cognate with either similar (14 filler items), or dissimilar (17 target items) lexical stress in Spanish and Dutch. The target words were counterbalanced for the presence of a written accent (marking where stress should be in Spanish). The phrases were presented individually one after the other on a screen, accompanied by a picture illustrating the meaning of the phrase. Only data from the target items was analysed. The items were presented in a different order at each time point.

2.2.2. Lexical stress training

Between T1 and T2, participants received lexical stress training, during which they were given written explanations about the 3 lexical stress rules governing lexical stress position in Spanish. Each rule was accompanied by a written example, which was also produced by a Spanish native in a video. Depending on the experimental condition (AV, AV-beat-perception, AV-beat-production, AV-metaphoric-perception, AV-metaphoric-production), the video showed the L1 speaker producing no gesture or either a beat or metaphoric gesture (horizontally visualising extended syllable duration) aligned with the stressed syllable of the target word and required different physical involvement levels by the participant (perceiving vs. also producing the gesture in training). After the example, participants were asked to imitate the speech of the speaker (and the hand movements, in the gesture production conditions), followed by another practice item and implicit feedback on this item via a prerecorded video of the Spanish native pronouncing the item (see Figure 1).

2.2.3. Musical aptitude task

After T2, participants performed musical aptitude and WM tasks. The musical aptitude task consisted of a modular version of the PROMS task which included short versions of the subsets that are most closely related to lexical stress: melody, rhythm, and accent. Subjects heard the same sound fragment twice followed by a comparison sound fragment and were asked to indicate whether the comparison fragment was the same or different from the first two fragments. Depending on the subset, participants were asked to focus on melody, rhythm, or accent.



Klemtoonregels in het Spaans

- **Regel 1: klinkers (a, i, o, u, e), -s of –n** Als een woord eindigt op een klinker of –s of –n, dan ligt de klemtoon op de één-na-laatste lettergreep.
- **Regel 2: medeklinkers (niet –s of –n)** Als een woord eindigt op een medeklinker (niet –s of –n), dan ligt de klemtoon op de laatste lettergreep.
- **Regel 3: accent** Als een woord een geschreven accent heeft, valt de klemtoon op die lettergreep.

Op de volgende slides worden de regels uitgelegd met voorbeelden.

1

Gebruik de muis om verder te gaan

Regel 1: klinkers, -s of -n

Als een woord eindigt op een klinker (a, i, o, u, e) of –s of –n, dan ligt de klemtoon op de één-na-laatste lettergreep.

2

La computadora violeta

Gebruik de muis om het fragment af te spelen. Je mag het fragment zo vaak afspelen als je zelf wilt. Probeer het ook hardop na te zeggen.

Gebruik de muis om verder te gaan

Oefen-item

Oefen hier nogmaals met de regel die je net geleerd hebt!

Je krijgt feedback op de volgende pagina.

El elefante gris

3

Gebruik de muis om verder te gaan

Feedback

Regel 1: klinkers (a, i, o, u, e), -s of –n Als een woord eindigt op een klinker of –s of –n, dan ligt de klemtoon op de één-na-laatste lettergreep.

4

El elefante gris

Gebruik de muis om het fragment af te spelen. Je mag het fragment zo vaak afspelen als je zelf wilt. Probeer het ook hardop na te zeggen.

Gebruik de muis om naar het volgende voorbeeld te gaan

Figure 1: *Lexical stress training slides. 1) the stress rules in Spanish, 2) an example, 3) a practice item, 4) feedback to the practice item in the AV condition.*

2.2.4. Working Memory task

The WM task was a backwards digit span task abstracted from the AWMA test battery [29]. Subjects heard digit spans and were asked to repeat them backwards. The task started with three easier practice items followed by six blocks of digit spans increasing in length (two to seven digits). If subjects produced four of the six spans in the block correctly, they went to the next block. If a subject made errors in more than three spans within one block the task ended. Subjects were scored on how many digit spans they repeated correctly. After the WM task, subjects did the final post-test (T3) and filled out a questionnaire about their language background and (formal) musical experience.

2.3. Procedure

In the lab, participants were seated behind a computer screen connected to a laptop that was controlled by the experimenter. The participant controlled the pace of the lexical stress training and musical aptitude tasks. The experimenter controlled the pace of the other tasks. The entire session was video- and audio-recorded and took 50-60 minutes, with a short break between the musical aptitude task and the WM task. The order of the different tasks was the same for all participants.

2.4. Data coding and analysis

Participants' speech productions at T1, T2, and T3 were transcribed and annotated in Praat [31]. Target items were coded (in a random order) as having correct Spanish lexical stress or not. 10% of the data was coded by a second rater, resulting in high inter-rater reliability, κ=.987, *p*<.001. The musical aptitude task resulted in a score for the subtests (i.e., melody (0 – 10), rhythm (0 – 8), and accent (0 – 10). For the WM task, participants received a score based on the final block they had reached (0-7).

A multilevel logistic regression analysis was conducted in R and Rstudio [32] using the *lme4* package [33]. The binomial outcome variable was Spanish lexical stress production (correct/incorrect). The main predictors included in the analysis were: *time of testing* (T1/T2/T3), *training condition* (AV/AV-B-perc/AV-B-prod/AV-M-perc/AV-M-prod), *stress rule* (i.e., whether the target word had a written accent or not), *melody score* (*M* = 5.09, *SD* = 1.76), *rhythm score* (*M* = 4.58, *SD* =

1.31), *accent score* (*M* = 5.14, *SD* = 1.31), and *WM score* (*M* = 4.43, *SD* = 1.03). Various subject characteristics were entered as control predictors, and *item* was included as a random factor. A stepwise forward procedure was used to add main effects and interactions to a baseline model. In the Results, we will focus primarily on the findings related to musicality.

3. Results

There was a significant main effect of *time of testing,* with subjects being less likely to correctly produce L2 lexical stress at T1 (21.85% correct) than at T2 (61.60% correct, *b* = 2.52, *SE* = .76, *p* < .001) and at T3 (60.7% correct, *b* = 3.35, *SE* = .75, *p* < .001), with no difference between T2 and T3. There was also a significant main effect of *training condition,* with participants in the AV-beat production condition being less likely to produce correct L2 lexical stress (42.89% correct) than participants in the AV (49.6% correct, *b* = -1.95, *SE* = .94, *p* = .037), AV-metaphoric perception (45.78% correct, *b* = -1.34, *SE* = .41, *p* = .011), and AV-metaphoric-production condition (53.92% correct, *b* = -1.52, *SE* = .46, *p* = .010). However, this effect reflects performances across all three testing times, including T1, pre-training. Hence, conclusions about the effectiveness of gesture and physical involvement in the training conditions cannot be drawn based on these results. The interaction between *time of testing* and *training condition* was non-significant. Hence, differences in performance on L2 lexical stress production between T1, T2, and T3 cannot be explained by the different training conditions (see Figure 2). Specifically, there was no significant difference in improvement from T1 to T2, from T2 to T3, and from T1 to T3 between training with or without gestures, between having beat or metaphoric gestures in training (i.e., gesture type), and between producing or perceiving gestures (i.e., physical involvement level).

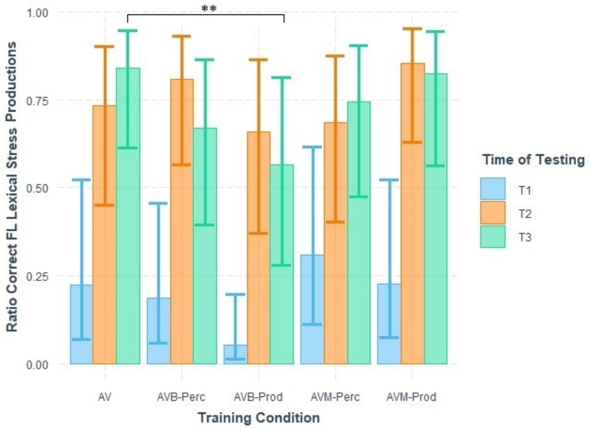


Figure 2. *Ratio of Correct L2 Lexical Stress Productions per Training Condition, Separated by Time of Testing*

The main effects of *WM*, *melody*, *rhythm*, and *accent scores* were all non-significant, indicating that subjects' WM capacity and musical aptitude did not directly influence L2 lexical stress productions (see Table 1). The two-way interactions between *time of testing* and *WM*, *melody*, *rhythm*, or *accent scores* were also non-significant, indicating that, across training conditions, subjects' improvements on producing L2 lexical stress were independent of their WM capacities or musical aptitudes. However, there was a significant three-way interaction between *time of testing*, *stress rule*, and *melody score* and between *time of testing*, *stress rule*, and *rhythm score* (see Table 2), where the relation between *time of testing* and *melody score* and between *time of testing* and *rhythm score* and their influence on producing correct L2 lexical stress varied between words with and without a written accent. In addition, several three-way interactions between *time of testing* and *training condition* on the one hand and *WM* or *musical aptitude scores* on the other hand were also significant, implying that WM capacity and musical aptitude influenced the effectiveness of the different training conditions over time, in different ways.

Table 1 *Mean (Standard Deviation) WM, Melody, Rhythm, and Accent scores split between incorrect and correct L2 lexical stress productions collapsed over Time of testing*

	L2 Lexical Stress Productions	
	Incorrect	Correct
	Mean (SD)	*Mean (SD)*
WM score	4.34 (1.03)	4.52 (1.02)
Melody score	4.97 (1.74)	5.24 (1.79)
Rhythm score	4.51 (1.32)	4.67 (1.29)
Accent score	5.08 (1.30)	5.22 (1.31)

Table 2 *Estimated Effects and coefficients for three-way interactions including individual differences*

Predictor	***b* estimate**	**Std. Error**	***z* value**	***p* value**
*T1 * AV-B-prod * WM score*	-2.60	1.20	-2.16	.031
*T1 * AV-M-prod * WM score*	2.54	1.25	2.03	.042
*T2 * Melody score * Stress rule*	1.50	.58	2.59	.010
*T2 * AV-B-prod * Melody score*	4.18	1.33	3.14	.002
*T3 * AV-B-prod * Melody score*	3.94	1.38	2.86	.004
*T3 * Rhythm score * Stress rule*	-.66	.24	-2.83	.005
*T2 * AV-B-prod * Rhythm score*	1.13	.36	3.15	.002
*T3 * AV-B-prod * Rhythm score*	2.00	.43	4.61	< .001
*T2 * AV-M-prod * Rhythm score*	1.24	.48	2.57	.001
*T3 * AV-M-prod * Rhythm score*	1.52	.54	2.83	.005
*T3 * AV-B-prod * Accent score*	-1.98	.69	-2.87	.004
*T1 * AV-M-prod * Accent score*	2.02	.91	2.22	.026

Note. Table only includes the significant three-way interactions from the regression model.

At T1, in the AV-B-prod condition, lower WM scores predicted better performance on L2 lexical stress production (which was also visible in the AV-B-perc condition), but in the AV-M-perc and AV-M-prod conditions, higher WM scores seemed to predict better performance. In the AV condition, WM scores did not predict performance. In addition, these results show that *melody aptitude scores* in combination with the AV-B-prod condition affect lexical stress production differently than the other conditions. The significant interaction between *time of testing*, *training condition*, and *rhythm score* suggests that the relationship between time of testing and rhythm aptitude varied across training conditions and that the relation between training condition and rhythm aptitude significantly varied across testing times. In the AV-B-prod and AV-M-prod condition, a higher rhythm aptitude predicted better performance on L2 lexical stress production at T2 and T3, whereas in the other training conditions this relation was less clear. The significant interaction between *time of testing*, *training condition*, and *accent score* (see Table 2) suggests that the relation between testing time and accent aptitude varied across training conditions. Whereas in the AV-B-prod condition lower accent scores predicted better performance, in the AV condition higher accent scores predicted better performance, and in the AV-B-perc condition accent scores did not predict performance on L2 lexical stress production. The significant three-way interaction also suggests that the relation between training condition and accent aptitude significantly varied across testing times.

4. Discussion & Conclusion

We aimed to disentangle the effect of gesture type (i.e., beat or metaphoric), physical involvement level (i.e., producing or perceiving gestures), and individual measures (WM capacity and musical aptitude) on the effectiveness of L2 lexical stress training. Firstly, the findings on time of testing showed that a short training session on a specific L2 prosodic contrast, in this case lexical stress, significantly improved L2 speakers' productions. An unanticipated finding was that there was no difference in outcome when comparing training with or without gestures or different gesture types. Likewise, there was no difference in effectiveness between producing or perceiving gestures during training. The lack of a significant interaction between *time of testing* and *training condition* showed that participants improved their performance after T1, but that it stayed the same between T2 and T3, irrespective of experimental condition. There were also no main effects of *WM*, *melody*, *rhythm*, and *accent scores.*

The overall lack of significance of main effects and two-way interactions may be due to a lack of statistical power: A power analysis [34] showed that the present study is limited as substantially more subjects were required per condition to reach statistical power. Nevertheless, there were several significant three-way interactions involving individual characteristics. Focussing on those involving musicality, first, melody scores significantly influenced differences between training conditions at T2 and T3, and not at T1, which is as expected, as, at T1, subjects had not yet received training. Moreover, at T2, subjects with a lower melody aptitude seemed to benefit most from producing or perceiving metaphoric gestures, while subjects with a higher melody aptitude seemed to benefit most from producing or perceiving beat gestures. At T3, these differences had disappeared, though producing beat gestures was still most beneficial for subjects with a higher melody aptitude, and perceiving metaphoric gestures was still most beneficial for subjects with a lower melody aptitude. Together, this suggests that the effectiveness of different gesture types might be related to L2 learners' melody aptitudes.

Second, the influence of rhythm scores on producing correct L2 lexical stress differed significantly before and after training. As expected, there were no significant differences between training conditions at T1, but at T2 and T3, subjects with a higher rhythm aptitude benefitted most from producing gestures in training, while subjects with a lower rhythm aptitude benefitted more from perceiving gestures. This implies that the effect of physical involvement level may depend on rhythm aptitude and that subjects with a lower rhythm aptitude benefit from less physical involvement. Third, the effect of accent scores on producing L2 lexical stress significantly differed before and after training. As expected, no significant differences between training conditions occurred at T1, but at T2, subjects with a lower accent aptitude benefitted more from all training conditions than subjects with a higher accent aptitude, except the metaphoric production condition, which was most effective for learners with a higher accent aptitude. Moreover, at T3, subjects with a lower accent aptitude benefitted most from producing and perceiving metaphoric gestures and producing beat gestures in training, while subjects with a higher accent aptitude benefitted most from training without gestures. These findings suggest that especially learners with a lower accent aptitude benefitted from training with gestures.

Although the influence of the different musical aptitude measures varies, the results indicate that subjects with a lower musical aptitude need more visual information compared to subjects with a higher musical aptitude in learning L2 prosody, but there is an important trade-off point: Whereas subjects with a lower melody aptitude benefitted more from metaphoric gestures, which are not acoustically related to speech but provide more information about the specific prosodic contrast compared to beat gestures, and subjects with a lower accent aptitude seemed to benefit more from gestural training in general, subjects with a lower rhythm aptitude did not benefit more from gesture production. Therefore, subjects with a lower musical aptitude seem to benefit from more visual information, as long as they can cognitively process this additional modality. Producing gestures might be too cognitively demanding for subjects with low musical aptitudes, who already must focus on the musical properties, and, therefore, perceiving gestures, which is less cognitively demanding, is more beneficial.

The results of this study have shown that, irrespective of training condition, subjects significantly improved their L2 lexical stress production after training. Moreover, we found several significant three-way interactions containing WM capacity or musical aptitude measures. Hence, the effectiveness of gesture type in combination with physical involvement level in L2 lexical stress training was significantly affected by these individual differences. Together, these findings suggest that, in L2 prosody acquisition, learner characteristics and fine-grained methodological choices (e.g., concerning gesture type and physical involvement level) together affect L2 prosody performance and should be carefully considered in future research.

5. Acknowledgments

We would like to thank Núria Domínguez for her help in creating the material.

6. References

[1] A. Kendon, "Some relationships between body motion and speech", In A.W. Siegman, and B. Pope, *Studies in dyadic communication*, New York, USA: Pergamon Press, pp. 177-210, 1972.

[2] D. McNeill, "So you think gestures are nonverbal?", *Psychological Review,* vol. 92, no. 3, pp. 350-371, 1985.

[3] S. Goldin-Meadow and C. Butcher, "Pointing toward two-word speech in young children". In S. Kita (Ed.), *Pointing: Where*

language, culture, and cognition meet, New York: Psychology Press, pp. 93-116, 2003.

[4] S. D. Kelly, T. McDevitt, and M. Esch, “Brief training with co-speech language lends a hand to word learning in a foreign language”. *Language and Cognitive Processes,* vol. 24, no. 2, pp. 313-334, 2009.

[5] M. Tellier, “The effect of gestures on second language memorisation by young children”, *Gesture,* vol. 8, no. 2, pp. 219-235, 2008.

[6] L. Quin-Allen, “The effects of emblematic gestures on the development and access of mental representations of French expressions”, *The Modern Language Journal,* vol.79, no. 4, pp. 521-529, 1995.

[7] D. McNeill, *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago press, 1992.

[8] T. Rietveld and V.J. van Heuven, *Algemene fonetiek,* Bussum, The Netherlands: Coutinho, 2009.

[9] T. M. Derwing and M. J. Munro, “Accent, intelligibility, and comprehensibility: Evidence from four L1s”, *Studies in Second Language Acquisition,* vol. 20, pp. 1-10, 1997.

[10] M. J. Munro and T. M. Derwing, “Foreign accent, comprehensibility, and intelligibility in the speech of second language learners”, *Language Learning,* vol. 45, pp. 73-97, 1995.

[11] M. J. Munro and T. M. Derwing, “Foreign accent, comprehensibility, and intelligibility in the speech of second language learners”, *Language Learning,* vol. 49, pp. 285-310, 1999.

[12] T. Derwing and M. Munro, “Preparation of teachers of English as a second language: Putting the NS/NNS debate in context”, In E. Llurda (Ed.), *Non-native language teachers. Perceptions, challenges and contributions to the profession,* Springer, pp. 179-191, 2005.

[13] E. Clark and A. Paran, “The employability of non-native-speaker teachers of EFL: A UK survey”, *System,* vol. 35, no. 4, pp. 407-430, 2007.

[14] B. Hendriks, F. Van Meurs and N. Hogervorst, “Effects of degree of accentedness in lecturers’ Dutch-English pronunciation on Dutch students’ attitudes and perceptions of comprehensibility”. *Dutch Journal of Applied Linguistics,* vol. 5, no. 1, 2016.

[15] D. P. Loehr, “Temporal, structural, and pragmatic synchrony between intonation and gesture”, *Laboratory Phonology,* vol. 3, no. 1, pp. 71-89, 2012.

[16] W. Pouw and J. A. Dixon, “Quantifying gesture-speech synchrony”, In A. Grimminger (Ed.), *Proceedings of the 6th Gesture and Speech in Interaction – GESPIN 6,* Paderborn: Universitaetsbibliothek Paderborn, pp. 75-80, 2019.

[17] P. Wagner, Z. Malisz and S. Kopp, “Gesture and speech in interaction: An overview”, *Speech Communication,* vol. 57, pp. 209-232, 2014.

[18] F. Baills, N. Suárez-González, S. González-Fuente and P. Prieto, “Observing and producing pitch gestures facilitates the learning of mandarin Chinese tones and words”, *Studies in Second Language Acquisition,* vol. 41, no. 1, pp. 33-58, 2019.

[19] B. Hannah, Y. Wang, A. Jongman, J. A. Sereno, J. Cao and Y. Nie, “Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers”, *Frontiers in Psychology*, vol. 8, 2017.

[20] D. Gluhareva and P. Prieto, “Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations”, *Language Teaching Research,* vol. 21, no. 5, pp. 609-631, 2017.

[21] J. Llanes-Coromina, P. Prieto and P. L. Rohrer, “Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task”, In Klessa K., Bachan J., Wagner A., Karpiński M., Śledziński D., *Proceedings of the 9th international conference on speech prosody,* Poznań, Poland: ISCA, pp. 498-502, 2018.

[22] K. Eng, B. Hannah, L. Leong and Y. Wang, “Can co-speech hand gestures facilitate learning of non-native tones?” *Proceedings of Meetings on Acoustics ICA2013,* vol. 19, no. 1, 2013.

[23] L. M. Morett and L. Y. Chang, “Emphasising sound and meaning: Pitch gestures enhance Mandarin lexicon tone acquisition”, *Language, Cognition, and Neuroscience,* vol. 30, no. 3, pp. 347-353, 2015.

[24] S. Kelly, A. Bailey and Y. Hirata, “Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts”, *Collabra: Psychology,* vol. 3, no. 1, 2017.

[25] A. Kendon, *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press, 2004.

[26] M. Wilson, “Six views of embodied cognition”, *Psychonomic Bulletin & Review,* vol. 9, no. 4, pp. 625-636, 2002.

[27] D. Özer and T. Göksun, “Gesture use and processing: A review on individual differences in cognitive resources”, *Frontiers in Psychology,* vol. 11, 2020.

[28] N. Kraus and B. Chandrasekaran, “Music training for the development of auditory skills”, *Nature Reviews Neuroscience,* vol. 11, no. 8, pp. 599-605, 2010.

[29] T. P. Alloway, *Automated working memory assessment: Manual,* London: Pearson Assessment, 2007.

[30] L. N. Law and M. Zentner, “Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills”, *PloS One,* vol. 7, no. 12, 2012.

[31] P. Broersma and D. Weenink, “Praat: Doing phonetics by computer” [Computer program]. Version 6.0.37, retrieved 17 April 2018 from http://www.praat.org/, 2018.

[32] RStudio Team, “RStudio: Integrated Development for R”. [Computer Program]. RStudio, PBC. Boston, MA. Version 1.3.1093, retrieved from: http://www.rstudio.com/, 2020.

[33] D. Bates, M. Maechler, B. Bolker and S. Walker, “lme4: Linear mixed-effects models using ‘Eigen’ and S4”. Version 1.1-26, https://cran.r-project.org/web/packages/lme4/index.html, 2020

[34] F. Faul, E. Erdfelder, A. Buchner and A.-G. Lang, “Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses”, *Behavior Research Methods,* vol. 41, pp. 1149-1160, 2009.