Melodic perception skills predict Catalan speakers' speech imitation abilities of unfamiliar languages

Peng Li[1], *Yuan Zhang*[1], *Xianqiang Fu*[1], *Florence Baills*[1], *Pilar Prieto*[2,1]

[1]Universitat Pompeu Fabra, Spain
[2]Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain
peng.li@upf.edu, yuan.zhang@upf.edu, xianqiang.fu01@estudiant.upf.edu, florence.baills@upf.edu, pilar.prieto@upf.edu

Abstract

Musical perception skills have been shown to influence second language speech production. Likewise, working memory may also affect nonnative speech production abilities. However, very few studies have assessed their respective role in speech imitation abilities. The present study thus investigates the predictive role of musical perception skills and working memory on speech imitation abilities of unfamiliar languages. Sixty-one adult Catalan speakers imitated twelve sentences in six languages that were unfamiliar to them. Participants' music perception skills were tested by four PROMS-S subsets, namely accent, melody, pitch, and rhythm, and their working memory was measured by a forward digit span test. A linear regression analysis revealed that melodic perception skills were the unique predictor among all four musical perception subtests and that working memory was not a significant predictor. Our findings show that melodic perception skills are key in predicting the capability in imitating unfamiliar speech, and thus may be important for learning foreign language pronunciation.

Index Terms: speech imitation, musical perception skills, working memory, melody

1. Introduction

In recent decades, increasing evidence from cognitive neuroscience has revealed the close relationship between music and language [1]. For instance, music and speech share a common perceptual system [2] and present neural overlapping in the processing of auditory stimuli [3].

Focusing on the role of individual differences in music skills, second language (L2) speech learning may be related to individual musical expertise and aptitude [4]. Regarding the role of musical expertise (i.e., musicianship) on speech processing and production skills, neuroscientific studies have revealed mixed results. Some studies found that musical expertise benefitted speech perception skills, such as the perception of pitch [5], tonal and segmental features [6], and speech segmentation [7], as well as imitation abilities, for example, with lexical tones [8]. Other studies, however, showed that musicality did not predict the perception of segmental features [9].

As for the role of musical aptitudes (both perceptual and productive), recent studies have also revealed their importance for L2 speech perception skills [10] and production skills [11]. For instance, pitch perception skills significantly predicted overall proficiency in L2 pronunciation [12] and the perception and production of lexical tones [13]; melodic and pitch perception skills also affected L2 intonation [14]. More broadly, musical perception and production abilities can both predict L2 word discrimination and pronunciation abilities [10], and higher musical perception skills can even entail better pronunciation of challenging L2 sounds [11].

Interestingly, the different subcomponents of musical aptitude (e.g., the sensitivity to rhythmic vs. melodic patterns) may play different predictive roles in L2 speech perception and production. First, on speech perception, rhythmic perception skills affected the perception of durational contrasts, whereas pitch perception skills did not [15]. Second, on speech production, musical timing perception skills predicted the imitation accuracy of challenging L2 phonemes (e.g., English /r/ and /l/ for Japanese speakers), whereas other perception skills (pitch, loudness, rhythm) were not significant predictors [16]. A recent study found that melodic and accent perception skills predicted imitation abilities of regional accent in one's native language (L1), but tempo and tuning could not [17]. Finally, another study showed that rhythm and pitch perception skills only weakly predicted the pronunciation of L2 prosodic features [18]. Therefore, these studies call for a more comprehensive assessment of musical abilities that involves different subcomponents.

Another line of research has attempted to explore the relationship between musical aptitude and speech imitation abilities by having participants imitate unfamiliar languages. Briefly, people with better musical aptitude performed better in the pronunciation of suprasegmental features [19] and achieved better speech intelligibility [20]. Moreover, some subcomponents of musical aptitude, like rhythmic production skills, could also predict the production of unfamiliar L2 stress-accent placement [21]. Finally, a positive association between musicality and speech imitation has been found in adults and pre-school children [22].

Importantly, previous studies on the relationship between musicality and speech production/imitation have also considered participants' working memory capacities a confounding factor (e.g., [10]) and have reported mixed results. Working memory is related to musical and phonological abilities and thus may affect L2 speech production [23]. For instance, working memory and singing abilities jointly predicted L2 speech imitation abilities, and singing abilities affected the working memory span [24]. A similar link between speech imitation, musical aptitude, and working memory has been found in pre-school children [22]. By contrast, there is little evidence of a direct relationship between working memory capacity and imitation performance of a language unfamiliar to the imitators [20].

In addition to its confounding role with musical abilities, the effect of working memory in L2 speech production has also produced mixed results. Focusing first on the positive results, working memory was positively associated with the development of L2 oral fluency [25], speech accuracy and complexity [26], and even overall L2 language proficiency, including speaking skills [27]. By contrast, some studies failed to find a correlation between working memory and L2 speech fluency [28], L2 sound production [15], or overall speech proficiency [12].

Given these mixed findings, it seems worth continuing to investigate the role of working memory when assessing potential differences in speech imitation abilities. Thus far, very few studies have examined the role of a variety of specific musical perception skills on speech imitation abilities. It might well be that the predictive role of musical perception skills on speech imitation skills may be domain specific. Because previous studies did not reach an agreement on which musical skill is the most relevant factor in speech imitation skills, it might well be that not all the aspects of musical aptitude equally predict speech imitation skills.

The present study aims to assess the relationship between four of the main subcomponents of musical perception skills (accent, melody, pitch, and rhythm), working memory, and speech imitation abilities.

2. Methods

2.1. Participants

We recruited 61 Catalan-speaking undergraduate students (54 female, M_{age} = 19.70 years, SD = 1.74) from the Department of Translation and Language Sciences at Universitat Pompeu Fabra. None of them reported having any auditory impairment nor exposure to the languages belonging to the language imitation task. The participants voluntarily participated in the study and signed a consent form allowing data collection and treatment for academic purposes.

2.2. Materials

The experiment consisted of three tasks: A forward digit span task to test working memory, a shortened perception test battery to assess music perception skills (accent, melody, pitch, and rhythm), and a speech imitation task with sentences in six languages that were unfamiliar to the participants to assess general speech imitation skills.

2.2.1. Profile of Music Perception Skills: PROMS-S

The PROMS (Profile of Music Perception Skills) is a musical test battery that measures an individual’s music perception skills [29]. For the present study, we chose four subsets (accent, melody, pitch, and rhythm) from PROMS-S as an index of the assessment of music perception skills, which proposes a shorter version of the full PROMS. Although it is a shortened version, PROMS-S showed a reliable test score and good internal consistency [30].

The accent subsection tested participants’ perceptual skills of emphasized musical notes with identical duration (rhythm). The melody subsection involved a series of eighth monophonic musical notes with constant rhythms to test the perceptual abilities of changes in tones. The pitch subsection asked participants to discriminate pure tones varying in pitch. The rhythm subsection consisted of a series of two-bar notes varying in duration but constant in intensity to test the temporal perception skills

2.2.2. Forward digit span task

Digit span is one of the measures that assess the phonological component of working memory [31]. It measures the “capacity-limited phonological short-term store” [32, p.27]. In order to keep the experiment to a reasonable length, we selected the classic forward digit span task. To this end, we followed Woods et al. [33] and developed an on-line test using the WinSCP software to collect data. The test program was created based on an on-line script [34], which was then modified by Navarro Pérez and Rohrer [35].

2.2.3. Speech imitation task

Twelve sentences in six languages (Chinese, Hebrew, Japanese, Russian, Turkish, and Vietnamese) were selected for the speech imitation task, two sentences per language. The syllable count of the sentences varied from 5 to 12 (M = 8.50, SD = 2.29). Each sentence was spoken by a native speaker of each language and recorded in a soundproof room.

2.3. Procedure

Upon arriving at the experimental room, each participant signed a consent form and was asked to sit in front of a computer. The equipment was sanitized every time after use for protection issues during the COVID19 pandemic, and proper social distancing was strictly required.

First, participants accessed the digit span test by clicking on a link to the test. A series of numbers appeared on the computer screen. The participants had to remember and type the digit series in the forward order in which it had appeared by clicking on the keypad on the screen. The first series consisted of a three-digit span. A correct response would increase the span by one digit, while two incorrect responses would reduce the length of the following list by one digit. The task would end at the 14^{th} trial regardless of how many digits had been presented. The system automatically recorded participants’ scores. The task lasted approximately five minutes.

Then, participants undertook the test battery for musical perception skills (i.e., the accent, melody, pitch, and rhythm subsets of PROMS-S). Each subtest consisted of 8-10 trials. Participants listened twice to the same stimulus for each trial, followed by a short interval and a comparison stimulus. Participants had to indicate whether the comparison stimulus differed from the initial one by choosing among five options: *definitely different*, *probably different*, *I don’t know*, *probably the same*, and *definitely the same*. Again, participants’ scores were automatically recorded. The musical test lasted approximately 15 minutes.

Finally, participants completed the speech imitation task on the on-line testing platform *Alchemer*. They listened to each model sentence twice consecutively and tried their best to imitate each sentence only once. Participants were audio-recorded throughout the task individually and the experimenters collected the recorded audio files upon completion of the task.

2.4. Data coding

We exported the scores of the musical perception skills (one score for each of the four subsets: accent, melody, pitch, and rhythm) and digit span from the experimental platforms (for more details on the scoring systems, refer to Law and Zentner [29] for the PROMS-S and to Woods et al. [33] for the forward digit span).

For the speech imitation task, we organized perceptual rating tasks to assess participant's imitation abilities. We recruited three native speakers for each of the six. The evaluation was established on a Likert scale from 1 to 9, where 1 meant very bad imitation and 9 meant a perfect imitation. Before performing the rating, the raters familiarize themselves with the rating criteria and procedure in an on-line training session. Raters were explicitly told to give a score based on their global impression of the pronunciation accuracy other than paying specific attention to individual sounds or suprasegmental features. When the raters stated that they had understood the task, they evaluated six example audios as an exercise. They then showed their rating scores and discussed any discrepancies in the rating until they reached an agreement. When all the raters showed clear consistencies in rating and understood that they should rate the general pronunciation, the raters performed the rating task on-line individually via the *Alchemer* platform. In total, six training sessions were administered, with each session targeting one language. Each rater spent about half an hour completing the task.

2.5. Statistical analyses

First, in order to check for the consistency between the three raters of each language, we calculated their interrater reliability scores. Intraclass Correlation Coefficients (ICC) were obtained using the *irr* package [36] based on six mean-rating (k = 3), consistency, two-way mixed-effects models. The results revealed acceptable to excellent ICCs across the three raters for the six languages (Table 1). Therefore, for each item, the rating scores of the three raters were averaged, yielding a mean speech imitation score per item.

Table 1: *Interrater reliability (ICC) between the three raters for each of the languages*

Language	ICC
Chinese	0.87
Hebrew	0.93
Japanese	0.88
Turkish	0.90
Russian	0.72
Vietnamese	0.93

Second, we built a Linear Mixed Model (LMM) to analyze the relationship between the speech imitation scores and the perceptual musical scores, and working memory scores, using the *lmer()* function from the *lme4* package [37] in the R program. The fixed factors were working memory score and the scores of the four subcomponents of PROMS-S, namely accent, melody, pitch and rhythm. All the variables were scaled and centered to zero before fitting the model. We first built a full model including all the possible random slopes for each of the two random intercepts: subject and item. The best fitting model was selected by the *buildmer()* function from the *buildmer* package [38]. The best model will be reported here. It involves a random slope of working memory for both subject and item. Significance of the main effects was assessed by the *lme4Test* package [39].

3. Results

Table 2 summarizes the descriptive data of all the variables on their original scale.

Table 2: *Means (M), standard deviations (SD), minimum (Min), and maximum values (Max) of the scores of musical perception tests (accent, melody, pitch, and rhythm), working memory, and speech imitation*

Measure	*M*	*SD*	*Min*	*Max*
Accent	4.76	1.67	2.00	9.50
Melody	4.80	1.75	1.50	8.50
Pitch	3.12	1.29	1.00	6.50
Rhythm	4.89	1.47	1.50	7.50
Working memory	7.01	1.36	4.77	11.17
Speech imitation	4.28	2.01	1.00	8.33

As for the LMM analysis (Table 3) with the speech imitation score as a response variable, only melody was found to be a significant predictor of L2 speech imitation abilities, namely, the higher the melody score, the more likely participants were of obtaining a higher score in L2 speech imitation (Fig. 1). The other three subcomponents of musical perception skills (i.e., accent, pitch, and rhythm) did not significantly predict the L2 speech imitation scores. Likewise, working memory did not significantly predict the L2 imitation abilities.

Table 3: *Results from the LMM predicting speech imitation score with accent score, melody score, pitch score, rhythm score, and working memory score (WM) as fixed effects and participants (subject) and items as random effects. Estimates (β) represent the change in speech imitation score resulting from a change in each fixed factor.*

Predictor	Fixed effects				Random effects	
					By subject	By item
	β	*SE*	*t*(711)	*p*	*SD*	*SD*
Intercept	0.01	0.16	0.03	.973	0.13	0.29
Accent	0.06	0.07	0.93	.354	-	-
Melody	**0.18**	**0.06**	**2.70**	**.007**	**-**	**-**
Pitch	-0.06	0.06	-1.09	.275	-	-
Rhythm	0.04	0.07	0.61	.539	-	-
WM	0.03	0.06	0.57	.569	0.02	0.01

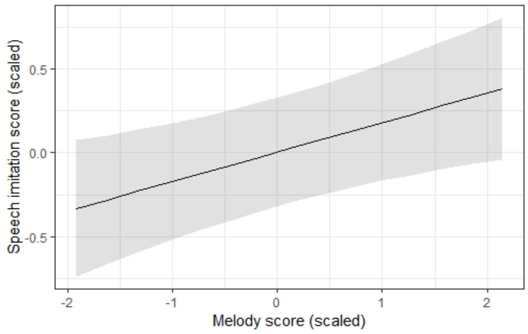

Fig.1: Linear regression analysis between the scaled melody score and the scaled speech imitation score. The gray band shows the 95% confidence interval.

4. Discussion

The present study investigated whether a set of musical perception skills and working memory capacity predicted the speech imitation skills of Catalan speakers. The results from the LMM analysis showed that: (a) only melodic perception skills showed a significant main effect on speech imitation abilities, and (b) working memory capacity was not a significant predictor of speech imitation performance.

Our results showed that participants with higher melodic perception skills were better at imitating speech in unfamiliar languages. Among the four subcomponents of musical perception skills, only melody significantly predicted speech imitation performance. This finding aligns with previous studies showing that melody perception skills may predict the accuracy of L2 nuclear accent production [14] and that melodic skills was predictive of L2 speech imitation abilities [17]. This finding thus conforms to the general belief that music and speech share a neural overlap. In addition to melody, Murljacic [17] found that accent perception skills also influenced the imitation of regional accent of English speakers. Because native speakers could understand the content of other regional varieties, when trying to imitate the regional accent, stress patterns may have been easier to detect. Therefore, accent perception skills were relevant for successful regional accent imitation. On the contrary, stress patterns may not be the first factor that imitators detect in unfamiliar languages. Therefore, accent may not play a role as important as melody in imitation unfamiliar languages. The current study thus expanded our knowledge on the relationship between speech and music by showing the predictive role of melody on imitation skills.

The skill transfer from music to language may be explained by the common features shared by musical melody and speech intonation [9]. That is, the participants with good melodic perception skills may have more easily perceived the suprasegmental features in an unfamiliar language and thus achieved a better performance in speech imitation. As the accuracy in suprasegmental features may affect native speakers' perception of accentedness [40], it might well be that good melodic abilities triggered good imitation abilities at the suprasegmental level, which resulted in higher speech imitation scores. Future research could separately assess the imitation accuracy at both the suprasegmental and segmental levels to explore which of the two linguistic components is responsible for better production skills.

Surprisingly, working memory capacity was not a significant predictor of speech imitation abilities. This finding suggests that when adults imitate relatively short sentences in a new language (mean N of syllables = 8.50), working memory may not be a crucial ability, which is in line with some previous studies [20]. In addition, the forward digit span test might be cursory to assess adults' working memory capacity. Therefore, the less sensitive measure may have underestimated the predictive value of working memory for speech imitation[1]. A future study might want to include more challenging tasks (such as backward digit span, non-word repetition span or even other complex tests, like reading span, speaking span, or equation span, see Wen [32] for a summary of working memory tests) to assess working memory capacities.

Beyond simple imitation skills, future studies could assess the predictive value of musical perception skills (and specifically melodic skills) for improving foreign language pronunciation abilities. Although accurate L2 speech production is challenging because languages differ in prosodic properties and specific sounds, our findings suggest that good melodic perceptual skills may trigger better pronunciation in a foreign language. Therefore, it can be hypothesized that when training L2 learners with challenging pronunciation, this individual factor may have impact on the learning outcome. Second, even though working memory capacity did not significantly predict productive phonology in the current study, it may be a mediator between music training and other aspects of L2 learning [12], such as the perceptual abilities of L2 speech [10]. Future studies could also include various working memory measures to check if there is a hierarchical structure in the effects of individual factors on L2 speech imitation/production.

5. Conclusions

To conclude, the current study highlights the links between musical perception skills (particularly melodic skills) and speech imitation skills of adults. This suggests that individual differences in musicality might be a crucial factor in L2 pronunciation training design.

6. Acknowledgements

This study is funded by the PGC2018-097007-B-l00 project awarded by the Ministerio de Ciencia, Innovación y Universidades (MCIU), Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER), Pla de Recerca de Catalunya: 2017 SGR_971, awarded by Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR). We sincerely thank Dr. Joan C. Mora (Universitat de Barcelona) and Dr. Sílvia Perpiñán for their comments on an earlier draft of the manuscript. Also, we thank Dr. Carmen Pérez-Vidal and Ms. Sharon Gutiérrez for helping us recruit participants. We would like to extend our thanks to Mr. Pedro Navarro Pérez and Mr. Patrick L. Rohrer for their help in creating the digit span task. Finally, we thank the students at the Department of Translation and Language Sciences, Universitat Pompeu Fabra, who voluntarily participated in this study.

7. References

[1] R. Milovanov and M. Tervaniemi, "The interplay between musical and linguistic aptitudes: A review," *Front. Psychol.*, vol. 2, no. NOV, pp. 1–6, 2011, doi: 10.3389/fpsyg.2011.00321.

[2] I. Peretz, D. Vuvan, M.-É. Lagrois, and J. L. Armony, "Neural overlap in processing music and speech," *Philos. Trans. B*, vol. 370, no. 20140090, 2015, doi: http://dx.doi.org/10.1098/rstb.2014.0090.

[3] M. Besson and D. Schön, "Comparison between language and music," *Annals of the New York Academy of Sciences*, vol. 930. pp. 232–258, 2001, doi: 10.1111/j.1749-6632.2001.tb05736.x.

[4] J. Chobert and M. Besson, "Musical expertise and second language learning," *Brain Sci.*, vol. 3, no. 2, pp. 923–940, 2013, doi: 10.3390/brainsci3020923.

[1] We thank an anonymous reviewer for the useful comment on the working memory measures.

[5] C. Marques, S. Moreno, S. L. Castro, and M. Besson, "Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence," *J. Cogn. Neurosci.*, vol. 19, no. 9, pp. 1453–1463, 2007, doi: https://doi.org/10.1162/jocn.2007.19.9.1453.

[6] C. Marie, F. Delogu, G. Lampis, M. O. Belardinelli, and M. Besson, "Influence of musical expertise on segmental and tonal processing in Mandarin Chinese," *J. Cogn. Neurosci.*, vol. 23, no. 10, pp. 2701–2715, 2011, doi: 10.1162/jocn.2010.21585.

[7] C. François, F. Jaillet, S. Takerkar, and D. Schön, "Faster sound stream segmentation in musicians than in nonmusicians," *PLoS ONE*, vol. 9, no. 7. 2014, doi: 10.1371/journal.pone.0101340.

[8] T. L. Gottfried, A. M. Staby, and C. J. Ziemer, "Musical experience and Mandarin tone discrimination and imitation," *J. Acoust. Soc. Am.*, vol. 115, no. 5, p. 2545, 2004.

[9] F. Delogu, G. Lampis, and M. O. Belardinelli, "From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception," *Eur. J. Cogn. Psychol.*, vol. 22, no. 1, pp. 46–61, 2010, doi: 10.1080/09541440802708136.

[10] L. R. Slevc and A. Miyake, "Individual differences in second-language proficiency: Does musical ability matter?," *Psychol. Sci.*, vol. 17, no. 8, pp. 675–681, 2006, doi: 10.1111/j.1467-9280.2006.01765.x.

[11] R. Milovanov, M. Huotilainen, V. Välimäki, P. A. A. Esquef, and M. Tervaniemi, "Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence," *Brain Res.*, vol. 1194, pp. 81–89, 2008, doi: 10.1016/j.brainres.2007.11.042.

[12] J. Posedel, L. Emery, B. Souza, and C. Fountain, "Pitch perception, working memory, and second-language phonological production," *Psychol. Music*, vol. 40, no. 4, pp. 508–517, 2012, doi: 10.1177/0305735611415145.

[13] M. Li and R. Dekeyser, "Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning," *Stud. Second Lang. Acquis.*, vol. 39, no. 4, pp. 593–620, 2017, doi: 10.1017/S0272263116000358.

[14] C. Yuan, S. González-Fuente, F. Baills, and P. Prieto, "Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers," *Stud. Second Lang. Acquis.*, vol. 41, no. 1, pp. 5–32, 2019, doi: 10.1017/S0272263117000316.

[15] P. Li, F. Baills, and P. Prieto, "Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts," *Stud. Second Lang. Acquis.*, vol. 42, no. 5, pp. 1015–1039, 2020, doi: 10.1017/S0272263120000054.

[16] M. Dolman and R. Spring, "To what extent does musical aptitude influence foreign language pronunciation skills? A multi-factorial analysis of Japanese learners of English," *World J. English Lang.*, vol. 4, no. 4, pp. 1–11, 2014, doi: 10.5430/wjel.v4n4p1.

[17] M. Murljacic, "Musical ability and accent imitation," Honors Scholar Theses, 2020.

[18] C. Zheng, K. Saito, and A. Tierney, "Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude," *Second Lang. Res.*, 2020, doi: 10.1177/0267658320978493.

[19] Z. Pei, Y. Wu, X. Xiang, and H. Qian, "The Effects of Musical Aptitude and Musical Training on Phonological Production in Foreign Languages," *English Lang. Teach.*, vol. 9, no. 6, p. 19, 2016, doi: 10.5539/elt.v9n6p19.

[20] F. Delogu and Y. Zheng, "Beneficial Effects of Musicality on the Development of Productive Phonology Skills in Second Language Acquisition," *Front. Neurosci.*, vol. 14, no. July, pp. 1–9, 2020, doi: 10.3389/fnins.2020.00618.

[21] N. Cason, M. Marmursztejn, M. D'Imperio, and D. Schön, "Rhythmic abilities correlate with L2 prosody imitation abilities in typologically different languages," *Lang. Speech*, vol. 63, no. 1, pp. 149–165, 2020, doi: 10.1177/0023830919826334.

[22] M. Christiner and S. M. Reiterer, "Early influence of musical abilities and working memory on speech imitation abilities: Study with pre-school children," *Brain Sci.*, vol. 8, no. 9, 2018, doi: 10.3390/brainsci8090169.

[23] G. Rota and S. M. Reiterer, "Cognitive aspects of pronunciation talent," in *Language Talent and Brain Activity*, G. Dogil and S. M. Reiterer, Eds. Berlin,Germany: Walter de Gruyter, 2009, pp. 67–96.

[24] M. Christiner and S. M. Reiterer, "Song and speech: Examining the link between singing talent and speech imitation ability," *Front. Psychol.*, vol. 4, no. NOV, pp. 1–11, 2013, doi: 10.3389/fpsyg.2013.00874.

[25] I. O'Brien, N. Segalowitz, B. Freed, and J. Collentine, "Phonological memory predicts second language oral fluency gains in adults," *Stud. Second Lang. Acquis.*, vol. 29, no. 4, pp. 557–581, 2007, doi: 10.1017/S027226310707043X.

[26] M. B. M. Fortkamp, "Working Memory Capacity and L2 Speech Production: An Exploratory Study," Universidade Federal de Santa Catarina,Brazil, 2000.

[27] J. Kormos and A. Sáfár, "Phonological short-term memory, working memory and foreign language performance in intensive language learning," *Bilingualism*, vol. 11, no. 2, pp. 261–271, 2008, doi: 10.1017/S1366728908003416.

[28] G. J. Mizera, "Working memory and L2 oral fluency," University of Pittsburgh, 2006.

[29] L. N. C. 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, doi: 10.1371/journal.pone.0052508.

[30] M. Zentner and H. Strauss, "Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS," *Ann. N. Y. Acad. Sci.*, vol. 1400, no. 1, pp. 33–45, 2017, doi: 10.1111/nyas.13410.

[31] A. Baddeley, "Working Memory," *Science (80-.).*, vol. 255, no. 5044, pp. 556–559, 1992.

[32] Z. Wen, *Working Memory and Second Language Learning: Towards an Integrated Approach*. Bristol: UK: Multilingual Matters, 2018.

[33] D. L. Woods *et al.*, "Improving digit span assessment of short-term verbal memory," *J. Clin. Exp. Neuropsychol.*, vol. 33, no. 1, pp. 101–111, 2011, doi: 10.1080/13803395.2010.493149.

[34] I. Eisenberg, Z. Enkavi, P. Bissett, V. Sochat, and R. Poldrack, "Digit-span." 2017, [Online]. Available: https://github.com/expfactory-experiments/digit-span.

[35] P. Navarro Pérez and P. L. Rohrer, "digit-span." 2020, [Online]. Available: https://github.com/pnavarro/digit-span.

[36] M. Gamer, J. Lemon, I. Fellows, and P. Singh, "irr: Various Coefficients of Interrater Reliability and Agreement version 0.84.1[software]." 2019, [Online]. Available: https://cran.r-project.org/package=irr.

[37] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear Mixed-Effects Models using {lme4}," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015, doi: 10.18637/jss.v067.i01.

[38] C. C. Voeten, "buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression." 2021, [Online]. Available: https://cran.r-project.org/package=buildmer.

[39] 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.

[40] K. Saito, P. Trofimovich, and T. Isaacs, "Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels," *Appl. Psycholinguist.*, vol. 37, no. 2, pp. 217–240, 2016, doi: 10.1017/S0142716414000502.