

The Predictive Roles of Musical Aptitude, Auditory Abilities, and Working Memory in L2 Speech Imitation: Differences Between Familiar and Unfamiliar Languages

Peng Li

Basque Center on Cognition, Brain and Language (BCBL), Spain;
Department of Linguistics and Scandinavian Studies,
University of Oslo, Norway
pli@bcbl.eu

Ioanna Ioannidou

Department of Linguistics and Scandinavian Studies,
University of Oslo, Norway
ioannai@student.iln.uio.no

Ilaria Marazzina

Department of Linguistics and Scandinavian Studies,
University of Oslo, Norway
ilariama@student.iln.uio.no

Paula Pericacho

Department of Linguistics and Scandinavian Studies,
University of Oslo, Norway
paulaper@student.iln.uio.no

Béibhinn Reardon

Department of Linguistics and Scandinavian Studies,
University of Oslo, Norway
bmreardo@student.iln.uio.no

Lu Xing

Department of Linguistics and Scandinavian Studies,
University of Oslo, Norway
luxi@student.iln.uio.no

Abstract

In learning a second language (L2), speech imitation ability is an important aptitude, which is subject to individual's familiarity to the target L2 and multiple cognitive factors. The present study investigates the impact of individuals' cognitive abilities on L2 speech imitation skills in a familiar (English) and an unfamiliar (Chinese) L2. Thirty-five L2 English speakers imitated English and Chinese short phrases and completed tests on musical aptitude (4 subsets: accent, melody, pitch, and rhythm), auditory processing abilities (3 subsets: duration, formant, and pitch), and working memory (2 subsets: forward and backward digit span). Their imitated speech was rated by native English and Chinese speakers. Globally, working memory is a stronger predictor for familiar L2 than for unfamiliar L2 and auditory processing abilities only predict the imitation abilities of familiar L2. Regarding specific components, musical melodic perception abilities, auditory pitch processing abilities, and forward digit span significantly predict L2 speech imitation regardless of language familiarity, but backward digit span predicts unfamiliar L2 imitation better than familiar L2. The results suggest that some cognitive factors (e.g., working memory) may be crucial at the first contact with a new language, whereas others (e.g., auditory processing) may be more relevant to more experienced learners.

Keywords: musical aptitude; auditory processing abilities; working memory; speech imitation ability; language familiarity.

Resumen

Para aprender una segunda lengua (L2), la capacidad de imitación del habla constituye una aptitud relevante, influida por el grado de familiaridad con la L2 y los factores cognitivos. El presente estudio examina el impacto de las habilidades cognitivas en la imitación del habla en una L2 familiar (inglés) y una no familiar (chino). Treinta y cinco hablantes de inglés como L2 imitaron frases breves en ambos idiomas y realizaron pruebas de aptitud musical (acento, melodía, tono y ritmo), procesamiento auditivo (duración, formantes y tono) y memoria de trabajo (retención de dígitos directos e inversos). Hablantes nativos de inglés y chino evaluaron su imitación. En general, la memoria de trabajo predice la imitación en la L2 familiar mejor que la no familiar, mientras que el procesamiento auditivo solo predice la L2 familiar. Específicamente, la percepción melódica musical, el procesamiento tonal auditivo y la retención de dígitos directos predicen la imitación en general, mientras que la retención inversa predijo la L2 no familiar mejor que la familiar. Los resultados sugieren que ciertos factores cognitivos (p. ej. memoria de trabajo) son claves en el primer contacto con una L2, mientras que otros (p. ej. procesamiento auditivo) cobran mayor relevancia con la experiencia.

Palabras clave: aptitud musical; habilidades de procesamiento auditivo; memoria de trabajo; habilidad de imitación del habla; familiaridad de lengua.

1. Introduction

The initial stage of learning a verbal language often involves imitation, especially for learning pronunciation and formulaic phrases (Berthier & Lambon Ralph, 2014; Ghazi-Saidi & Ansaldo, 2017), although imitation is not the only way of language acquisition (Lightbown & Spada, 2013). Speech imitation ability constitutes a vital aspect of second language (L2) acquisition. Skillful imitators show larger articulation space which grants them more access to a larger phonetic repertoire for novel sound learning compared to less skilled imitators (Reiterer et al., 2013). Moreover, speech imitation ability is not an isolated language aptitude, but rather affected by multiple cognitive factors, among which, musical aptitude and working memory have been identified as the main predictors of imitation skills in *unfamiliar* speech (Christiner et al., 2018; Christiner & Reiterer, 2018). Moreover, recently research has emphasized that auditory processing abilities may account for successful L2 speech learning as well (Saito, 2023). However, it is not clear whether auditory processing predicts speech imitation skills because most of the previous research focused on its role in actual L2 learning outcomes with spontaneous speech measures (e.g., Zheng et al., 2022). More importantly, individuals may perform differently in imitating a familiar L2 than an unfamiliar L2 (Christiner & Reiterer, 2013), which resemble novice learners and experienced learners. It is therefore necessary to consider the interplay between cognitive factors and language familiarity in predicting imitation skills. To draw a complete picture, the present study examines the different predictive roles of musical aptitude, auditory processing abilities and working memory in familiar and unfamiliar L2.

1.1. *The role of musical aptitude in L2 speech perception and production*

Cognitive neuroscience research has revealed a compelling relationship between music and language (Milovanov & Tervaniemi, 2011; Peretz et al., 2015). Music and speech rely on similar acoustic cues and neural mechanisms (Besson & Schön, 2001; Peretz et al., 2015). Accordingly, individuals with musical training have different neurological encoding of speech (Patel, 2011). To explain the subcortical encoding of speech observed in musicians, Patel (2011) proposed the OPERA (**O**verlap, **P**recision, **E**motion, **R**epetition, and **A**ttention) hypothesis which claims that the brain networks that process acoustic information for both speech and music may be triggered with high precision for speech processing. As a result, individuals with a better musical aptitude have better linguistic performance than those who are less talented in music.

Musical aptitude is an inherent potential that can facilitate music learning (Law & Zentner, 2012). Musical aptitude is associated with L2 acquisition or performance. Specifically, musical perception aptitude contributes to phonological awareness as measured by accent-faking tasks (Coumel et al., 2019, 2023), knowledge of receptive and productive L2 phonology (Slevc & Miyake, 2006), and L2 speech imitation intelligibility (Delogu & Zheng, 2020). Moreover, musical aptitude is positively correlated with the perceived accentedness of speech imitated by speakers who do not speak those languages. For instance, adults with a high singing talent can better imitate unintelligible and unfamiliar speech sounds (Christiner & Reiterer, 2013).

Nevertheless, there is no agreement on the specific component(s) of musical aptitude that predict L2 speech perception and production. These components may play different roles in specific phonetic and prosodic domains, among which rhythm and pitch have been largely discussed. Rhythmic perception abilities correlate with the perception of L2 speech rhythm (Boll-Avetisyan et al., 2017); durational contrasts (Li et al., 2020), and the imitation accuracy of unfamiliar languages (Christiner & Reiterer, 2013). Pitch perception abilities are mainly associated with L2 tones (Bowles et al., 2016; Christiner et al., 2022; Li & Dekeyser, 2017), L2 intonation production (Yuan et al., 2019), and global pronunciation proficiency (Posedel et al., 2012). Compared to rhythm and pitch, results are less conclusive for the other components of musical aptitude. For example, melodic skills positively correlate with L2 lexical tone perception (Delogu et al., 2010) and intonation production (Yuan et al., 2019), but not with vowel formants (Jekiel & Malarski, 2021). Musical perception abilities in accent and/or melody predict the imitation of unfamiliar regional accents (Murljadic, 2020) and unfamiliar languages (Li et al., 2024), as well as L2 lexical stress perception (Li & Xi, 2024).

In summary, previous studies have revealed inconsistent findings on the role of musical aptitude components in L2 speech. More importantly, the role of musical aptitude may differ in imitating familiar and unfamiliar L2s. However, to the best of our knowledge, only Christiner and Reiterer's (2013) study directly compared how musical aptitude affects imitation skills in familiar and unfamiliar L2s. They had L1 German speakers imitate English (familiar) and Hindi (unfamiliar) and found that musical perception abilities were only relevant for unfamiliar L2, with rhythm and voice quality perception abilities being positive predictors. However, the participants in Christiner and Reiterer's (2013) study were singers with professional musical training. Therefore, it is unclear whether musical expertise plays a confounding role with musical aptitude. Hence, a design with non-musicians imitating both familiar and unfamiliar L2s is needed to advance this topic.

1.2. The role of domain-general auditory processing abilities in L2 speech perception and production

Beyond musical perception abilities, general hearing abilities or auditory processing abilities are fundamental constructs for speech sound processing (Saito, 2023). Auditory processing abilities refer to “a set of lower-order abilities related to precisely perceiving individual dimensions of acoustic information” (Saito, 2023, p. 525). Better auditory abilities entail higher L2 lexical proficiency (Saito et al., 2022), better speech perception (Saito et al., 2022), and production accuracy (Saito et al., 2020), although they show limited effects on the improvement in perception after studying abroad (Sun et al., 2021) and L2 speech fluency (Saito et al., 2020).

Some studies investigated which specific auditory traits precisely predict L2 speech performance. The tests include pitch, formants, duration, and intensity. The predictive role of the specific component of auditory processing abilities appears to be subjective to the linguistic domains. For example, formant discrimination abilities account for the successful production of challenging L2 sounds (Saito et al., 2022) and segmental accuracy improvements after a study abroad program (Saito et al., 2020), but not suprasegmental learning or fluency measures (Saito et al., 2020; Zheng et al., 2022). Similarly, while duration discrimination predicts the segmental and lexical stress accuracy, it fails to predict intonation accuracy and optimal speed (Zheng et al., 2022). Finally, pitch discrimination abilities positively affect L2 segmental and intonation accuracy, but not word stress or optimal speed (Zheng et al., 2022).

To sum up, auditory processing is essential for L2 speech perception and production. However, since most of previous research dealt with proficient L2 speakers, it remains unclear how individual differences interact with factors related to language familiarity. For instance, would auditory processing abilities equally predict the imitation performance of familiar L2 and unfamiliar L2? The latter one resembles a learner’s initial contact with a novel language, which provides insights into L2 teaching and training. Moreover, it is unclear which specific components are the most relevant predictors. Again, language familiarity may also play a role in this regard.

1.3. The role of working memory in L2 speech perception and production

While human communication relies heavily on auditory cues, the processing of such cues involves a series of executive functions, among which, working memory accounts for the updating and mental manipulation of information (Darcy et al.,

2016; Mora & Darcy, 2023). Working memory refers to the temporary storage and simultaneous manipulation of information during cognitive processes, providing interfaces between perception, long-term memory, and actions (Baddeley, 1992). According to the Multicomponent Model of Working Memory (Baddeley et al., 2021), speech sounds and musical stimuli are first retrieved through hearing resources and stored temporarily in the phonological loop, which is then supported by the executive function through the episodic buffer. Accordingly, better phonological short-term working memory allows more speech signals to be stored for processing, which enhances language comprehension, acquisition, and production (Baddeley, 2003; Baddeley et al., 1988, 1998).

L2 speech perception and production require high working memory loads (Alia-García & Mora, 2008; Fortkamp, 2000; O'Brien et al., 2007; Trude & Tokowicz, 2011), especially at the beginner level (O'Brien et al., 2006) or during initial contact with the new language, as reflected by the processing and imitation of unfamiliar languages (Christiner & Reiterer, 2013, 2015, 2018). This highlights the crucial role of working memory in phonological awareness and the retention of speech sound sequences. By contrast, several studies have reported limited predictive effects of working memory on imitation accuracy in unfamiliar languages (Coumel et al., 2019; Li et al., 2020; Li et al., 2022). Similar findings were also revealed for the productive abilities of a familiar L2, such as fluency (Mizera, 2006), and overall pronunciation accuracy (Cho, 2018; Posedel et al., 2012).

In short, previous studies have revealed contradictory evidence regarding the role of working memory in L2 speech production. When considering the role of language familiarity, more evidence is needed to clarify whether the role of working memory differs in predicting the imitation of familiar and unfamiliar L2s. Taken together, it seems relevant to investigate how the three cognitive factors—musical aptitude, auditory processing abilities, and working memory—predict L2 speech imitation skills while considering the role of familiarity. In the present study, language familiarity refers to whether an imitator has prior knowledge in the target L2 they are imitating.

1.4. The present study

This study aimed to explore the influence and interactions of multiple individual factors (musical aptitude, auditory processing, and working memory) on the imitation skills in a familiar L2 (with previous language learning experience) and an unfamiliar L2 (with no previous language learning experience). The familiar L2 was English and the unfamiliar L2 was Chinese.

We address the following research questions and hypotheses:

RQ1: Which cognitive factors predict L2 speech imitation skills, and does language familiarity affect the predictive values of cognitive factors?

H1a: Musical aptitude, auditory processing abilities, and Working memory would predict L2 speech imitation skills.

H1b: Imitators can retrieve previous knowledge when imitating a familiar L2, while when imitating an unfamiliar L2, the lack of previous knowledge would lead imitators to seek more cognitive resources to complete the imitation. We hypothesized that cognitive factors would be more predictive in unfamiliar L2 than in familiar L2.

RQ2: In each of the cognitive factors, which components would be the most predictive variables, and would language familiarity play a role?

H2: It will remain exploratory as to which components of each cognitive ability would be the most predictive, but it is reasonable to hypothesize that the predictive value of the specific components would be more robust in unfamiliar L2 than in familiar L2.

2. Method

2.1. Participants

The participants ($N = 35$, 20 females, aged 19 – 50 years, $M = 29.14$, $SD = 8.58$) were adult, non-native speakers of the familiar L2, English, with no prior knowledge of the unfamiliar L2, Chinese. The sample size was determined based on two factors. First, to assess imitation abilities, the participants would imitate 18 phrases (see Section 2.2 for details), and each phrase would be rated by human raters (see Section 2.4 for details). To avoid raters' fatigue and ensure rating validity, the rating session should be controlled to a reasonable duration. Following previous research where the rating sessions usually last between one to two hours (Sun et al., 2021; Zheng et al., 2022), we decided to recruit around 30-40 participants so that the rating could be completed within two hours. Second, we would use linear mixed effects models to conduct trial-based analyses which consider variances on both participant ($N = 35$) and item (18 phrases) levels (see Section 2.5 for details). This approach can even ensure good statistical reliability in extremely small- N designs (see Wiley & Rapp, 2019 for a five-participant study). Therefore, we con-

sider that our current sample size is reasonable in terms of logistic feasibility and statistical validity.

To ensure that our findings are generalizable cross-linguistically, following Christiner et al. (2018), we recruited participants from diverse linguistic backgrounds whose L1s include Arabic (1), Dutch (2), German (3), Greek (3), Italian (2), Kurdish (1), Nepali (2), Newari (1), Norwegian (11), Polish (2), Serbian (1), Spanish (3), Swahili (1), Swedish (1), and Turkish (1). Moreover, the predictive roles of individual factors such as musical aptitude on imitation abilities are different between lexical tone language speakers and non-tone language speakers (Li et al., 2024). Because the current study was conducted in a European university, we decided to focus on non-tone language speakers due to logistic feasibility. As a result, any individuals without tone language knowledge but with fluent English knowledge who responded to our call for participants were considered eligible in terms of linguistic background.

All the participants were enrolled in MA courses with English being the instruction language, which ensured that the participants had to use English frequently and were familiar with English. Because the imitation task requires speaking and listening abilities, we asked the participants to self-estimate their English proficiency in speaking and listening from 1 (very poor) to 7 (native/near native). Then, following Li et al. (2020), we aggregated and converted the two scores to a 0-1 scale, calculated as “Proficiency = $0.5 \times \text{speaking score} / 7 + 0.5 \times \text{listening score} / 7$ ”. Overall, the participants reported an advanced level of listening and oral proficiency in English ($M = 0.86$, $SD = 0.14$), which was further validated through their verbal communications in English with the experimenters.

Finally, none of the participants had received professional musical education, except for regular musical classes at school, nor did they consider themselves (semi-) professional musicians. All participants reported normal hearing with no documented cognitive or speech disorders. Each participant gave written consent, which allowed the researchers to process their personal data.

2.2. Materials

The experiment had four tests: L2 speech imitation, musical aptitude (accent, melody, pitch, and rhythm), auditory processing (duration, formant, and pitch), and working memory (forward and backward digit span task). The materials used for each part are described in the following subsections.

2.2.1. L2 speech imitation task stimuli

A total of 18 phrases, nine in English and nine in Chinese, were designed for the L2 speech imitation task and recorded by one native speaker per language (Appendix A). Syllable length was decided based on the average number (7 ± 2) of discrete items commonly recalled during working memory tests (Miller, 1994). Thus, the English phrases were 7, 9, or 11 syllables long, while the Chinese phrases were 4, 6, or 8 syllables long. We decided to set the syllable length of the unfamiliar L2 to be lower than that of the familiar L2, after pilot trials revealed that syllable lengths beyond eight in Chinese were often too challenging for novice imitators. The materials can be found on OSF via this view-only link: https://osf.io/g72y6/files/osfstorage?view_only=d84b04b30257466bb51142d301393948.

To make the task challenging, the phrases included difficult phonetic features for non-native speakers of English and Chinese. For instance, the English / δ - θ / and / i - $ɪ$ / contrasts, connected speech, and vowel reduction; the Chinese alveolo-palatal sibilants / $t\epsilon$, $t\epsilon^h$ - ϵ /, approximants (“apical vowels”) / $ɪ$, $ɹ$ / and the like (Chen et al., 2013).

2.2.2. PROMS-S test

We chose four subtests (accent, melody, pitch, and rhythm) from the short Profile of Music Perception Skills (PROMS-S) (Law & Zentner, 2012) to assess participants’ musical perception abilities, following previous research (Li et al., 2022). Zentner and Strauss (2017) confirmed that PROMS-S has robust internal consistency (McDonald’s $\omega = 0.92$), test-retest reliability ($r = 0.88$), and good discriminant validity ($r = 0.49$, moderate correlation between subsets). Therefore, PROMS-S is a standard test battery and has been used in many previous similar studies (e.g., Li et al., 2020; Li et al., 2022, 2024; Yuan et al., 2019).

The stimuli of the accent subtest were musical notes of equal duration but differed in intensity to test the participants’ discriminative abilities of the emphasized notes. The melody subtest assessed the perception of pitch changes using a series of monophonic eighth notes played with a constant rhythm. The pitch subtest consisted of pure tones varying in pitch to test the participants’ perceptual abilities to pitch changes. Finally, the rhythm subtests tested the participants’ perception abilities in temporal changes using a series of two-bar notes with constant intensity but varying duration. For the full test materials, refer to <https://webapp.uibk.ac.at/psychologie/musiquote/index.php/868468/lang-en>. In our experiment, each participant was assigned a unique participant code for identification purposes. To take this test anonymously, one can write “visitor” as the participant code.

2.2.3. Auditory processing test

We selected three subsets from the Auditory Processing-Discrimination Test Battery (Mora-Plaza et al., 2022) to measure the degree to which participants could perceive subtle changes in duration, formant, and pitch. The test battery is a recently developed platform which has been widely used in a series of works (most of the literature reviewed in Section 1.2 used this platform) and can tap into different domains of auditory processing abilities (Saito & Tierney, 2024). A recent test-retest analysis on this test battery showed an overall intraclass correlation coefficient of 0.689, which is considered fair to excellent by the authors (Saito & Tierney, 2024). The testing materials can be found at https://drive.google.com/drive/folders/1zpqAeeb5y9dlonpF_P-GejJwg7cBrvEs.

2.2.4. Online test platform for working memory

The working memory test involved two tasks: a forward digit span and a backward digit span. We used an online testing platform for this task (Denk, 2023), which followed the classic testing paradigm of digit span. The test can be accessed at <https://tools.timodenk.com/digit-span-test>.

2.3. Procedure

Participants were tested in a quiet room. After signing the consent forms, they performed the tasks according to the experimenter's instructions.

In the L2 speech imitation task, participants practiced the task using two exercise phrases, that were not part of the 18 test phrases. They then completed the test with the 18 target phrases via pre-timed presentations played on a computer. Participants could hear each phrase via a headset twice and imitate it once while being recorded by the Recorder software on the experimental computer. Both the order of the languages and phrases within each language set were randomized for each participant.

The participants then underwent PROMS-S tests in accent, melody, pitch, and rhythm, with each subtest containing 8-10 trials. In each trial, the participants first listened twice to the referent stimulus, followed by a comparison stimulus. They had to indicate whether the comparison differed from the referent and choose one answer from five options: *definitely different*, *probably different*, *I don't know*, *probably the same*, and *definitely the same*.

Next, the participants performed the auditory processing test using an adaptive, forced-choice procedure. Participants heard a sequence of three stimuli, separated by a 0.5s interval, and had to select whether the first or the third stimulus differed from the second by clicking the numbers “1” or “3” on screen. Initially, the task became more difficult; that is, the differences between the stimuli became subtle when participants gave three correct responses. Conversely, the task became easier after an incorrect response. The tests were stopped after either 70 trials had been completed or after eight reversals had occurred.

Finally, the participants performed the working memory test. In the digit span tests, digits appeared on the screen individually. To remove interference from auditory input, we unchecked the “enable sound” box so that the digits appeared on the screen in silence. In the forward mode, participants had to recall the digits by typing them on a keyboard in forward order. In the backward mode, participants had to type in numbers inverting the order in which they had seen them. For both tasks, we used a default speed of 1,000ms per digit, and the lowest possible span was four digits. A correct response would increase the span by one digit, whereas an incorrect response would end the test.

2.4. Data coding

2.4.1 L2 speech imitation rating

In total, the participants produced 623 imitation phrases (35 participants × 18 phrases – 7 missing trials). Three native speakers of each language rated each sentence on a scale ranging from 1 (very bad imitation) to 9 (perfect imitation). The raters were MA or PhD students in linguistic studies at a public university. All had received training in phonetic studies. The English raters (2 females and 1 male) were all British English speakers to match the accent of the speaker who provided the model speech for the imitation task. The Chinese raters were all from Mainland China (2 females and 1 male) with Mandarin Chinese being their native language, which also matched the linguistic profile of the model speaker of the imitation task. Before rating, raters received training on applying the rating scale, practicing 10 of the exercise phrases imitated by the participants, which were not included in the final analyses (see Section 2.3). During the training phase, raters could discuss the rating criteria until they reached agreement. In the formal rating session, the audio clips were played in a randomized order, and raters worked individually without discussion.

2.4.2. Music aptitude scores

The PROMS-S test battery automatically generated four scores for each participant: accent, melody, musical pitch, and rhythm. For each trial, a correct answer with “probably” was awarded 1 point and with “definitely” was awarded 2 points. A wrong answer or “I don’t know” received 0 point. The sum of the trials was the score of each subset.

2.4.3. Auditory processing scores

The auditory processing test battery automatically generated three scores: duration, formant, and auditory pitch. Lower scores indicated more precise auditory processing abilities in each subset.

2.4.4. Working memory score

Working memory tests yielded two scores: forward and backward digit spans, which were the maximum number of digits that a participant could correctly recall in each mode.

2.5. Statistical Analysis

Data analysis was conducted using R (R Core Team, 2014) and the following packages: *irr* (Gamer et al., 2019), *lme4* (Bates et al., 2015), and *buildmer* (Voeten, 2021). The raw data and statistical analysis can be found on OSF via this view-only link: https://osf.io/g72y6/files/osfstorage?view_only=d84b04b30257466bb51142d301393948. The *irr* package was selected following Hallgren’s (2012) recommendations for calculating inter-rater reliability with R software. This package provides a wide range of coefficients and their 95% confidence intervals (95% CI) which are critical for validating the reliability and replicability of human rating. The *lme4* package is one of the most used R packages for performing hierarchical regression models as it supports complex random structures (Bates et al., 2015), and is especially compatible with model selection tools like *buildmer*. Finally, we used *buildmer* to automatically process stepwise model selection and optimize both fixed and random effects structures (Voeten, 2021). This approach can refine the model for RQ1 and conduct exploratory analyses for RQ2.

Inter-rater reliability was calculated using the *irr* package’s Intra-Class Correlation analysis function. The raters showed good levels of reliability in English, ICC = 0.88, 95% CI [0.86, 0.91], and Chinese, ICC = 0.88, 95% CI [0.86, 0.90]. Therefore,

each participant's imitation score per item was calculated by averaging the three rating scores, which was labelled as "imitation score".

To test H1, we averaged the scores of the components of each cognitive variable to give an overall score, so that we obtained working memory, auditory, and music scores. We then submitted the three scores as independent variables to a linear mixed effects model using *lme4* package (Model 1), with imitation score as the dependent variable. Model 1 also added language (two levels: English vs. Chinese) as a fixed effect and the two-way interaction between language and each of the three cognitive scores. To test H2, we built three models using *lme4* package, each targeting one cognitive variable (Model 2: musical aptitude; Model 3: auditory processing; Model 4: working memory). All the numeric variables were z-score transformed before submitting to the models.

For each model, we first created a maximal model including all independent variables and their interactions with language (Chinese vs. English) and two random intercepts, participant and item, with all possible random slopes. For Model 1 (H1), we used *buildmer* to find the best regression model that could converge without eliminating any fixed effect or any of the interactions. By contrast, for Models 2-4 (H2), because RQ2 is partially exploratory, we used *buildmer* to find the optimal random structure and applied a stepwise backward elimination of the likelihood-ratio test to determine the fixed factors to be included in the final models.

3. Results

Table 1 outlines the descriptive statistics for the independent and dependent variables.

Table 1

Mean (M), standard deviations (SD), minimum (min), and maximum (max) values of the variables

	M	SD	Min	Max
Imitation score				
English	4.96	1.86	1.00	9.00
Chinese	3.14	1.68	1.00	8.33
Music score				
Accent	4.50	1.68	1.00	8.00
Melody	5.13	1.60	2.50	8.50
Pitch	5.30	1.68	1.00	9.50
Rhythm	3.69	1.65	1.50	7.00
Auditory processing score				
Duration	16.76	11.73	4.00	51.83
Formant	31.33	15.66	4.66	79.50
Pitch	13.37	8.75	4.16	44.33
Working memory score				
Forward digit span	5.43	1.33	4.00	8.00
Backward digit span	5.14	1.24	4.00	9.00

3.1. RQ1: Which cognitive factors predict the L2 speech imitation skills and does language familiarity affect the predictive values of the cognitive factors?

The results of Model 1 (Table 2) can be interpreted as follows. First, musical aptitude score was not a significant predictor of imitation score and did not interact with language. Second, auditory score significantly predicted the English imitation score, but not the Chinese imitation score. As lower auditory score means better processing abilities, the negative coefficient means that better auditory precision predicted higher English imitation score. Third, working memory significantly predicted both Chinese and English imitation skills, and the predictive value was stronger

for Chinese than for English. In sum, only auditory processing and working memory scores were significant predictors of imitation scores, and their predictive values were subject to language.

Table 2

Mixed effects model with imitation score predicted by cognitive scores (music aptitude, auditory processing, and working memory), language (Chinese vs. English), and the two-way interactions between the cognitive scores and language

	Fixed effects			Random effects	
	β	95% CI	<i>P</i>	by	by
				participant	item
				SD	SD
(Intercept = Chinese)	-0.46	[-0.72, -0.20]	.002	0.10	0.12
Music aptitude	0.02	[-0.13, 0.17]	.791		0.01
Auditory processing	0.01	[-0.12, 0.14]	.905		
Working memory (WM)	0.31	[0.18, 0.45]	<.001		
Language [English]	0.90	[0.47, 1.34]	<.001	0.57	
Music × Language [English]	-0.10	[-0.45, 0.25]	.578		
Auditory × Language [English]	-0.35	[-0.66, -0.04]	.035		
WM × Language [English]	-0.37	[-0.70, -0.04]	.038		

Model 1 formula: imitation ~ (music + auditory + working memory) * language + (language | participant) + (music | item).
Marginal $R^2 = 0.32$; Conditional $R^2 = 0.78$.

3.2 RQ2: In each of the cognitive factors, which components would be the most predictive variables, and would language familiarity play a role?

3.2.1 Musical perception abilities and imitation score

After backward elimination, Model 2 included only melody score and language as the main effects without interaction. Therefore, melody was the only significant musical component that predicted language imitation ability, regardless of language familiarity (Table 3). The positive coefficient means that individuals with better melodic perception abilities can imitate an L2 better regardless of familiarity. Accent, musical pitch, and rhythm were not significant predictors. The main effect of lan-

guage was not surprising, as an imitator is expected to imitate a familiar L2 much better than an unfamiliar L2.

3.2.2 Auditory processing abilities and language imitation skills

Model 3 (Table 3) only involved the significant main effects of auditory pitch and language. Therefore, pitch was the only significant auditory component to predict language imitation. As low auditory processing score indicated accurate auditory processing abilities, the negative coefficient means that individuals with more precise pitch processing abilities showed better L2 imitation score, regardless of language familiarity. Duration and formant were not significant predictors.

3.2.3 Working memory and language imitation skills

Model 4 involved the main effects of forward digit span, backward digit span, language, and two-way interaction between backward digit span and language (Table 3). The results showed that L2 imitation skills were predicted by the forward digit span regardless of L2 familiarity. However, although the backward digit span could predict L2 imitation skills in general, its predictive value was stronger for unfamiliar L2 than for familiar L2.

Table 3

Mixed effects models with imitation score predicted by the component scores of musical perception, auditory processing, working memory, and language, as well as possible two-way interactions

	Fixed effects			Random effects	
	β	95% CI	<i>P</i>	by participant SD	by item SD
Model 2: Music perception and imitation score					
(Intercept = Chinese)	-0.47	[-0.74, -0.20]	.002	0.40	0.35
Melody	0.19	[0.05, 0.32]	.012		0.08
Language [English]	0.92	[0.47, 1.37]	<.001	0.92	
Model 3: Auditory processing and imitation score					
(Intercept = Chinese)	-0.46	[-0.74, -0.19]	.003	0.44	0.35
Auditory pitch	-0.16	[-0.29, -0.02]	.030		
Language [English]	0.90	[0.45, 1.36]	<.001	0.91	

Model 4: Working memory and imitation score

(Intercept = Chinese)	-0.46	[-0.72, -0.21]	.002	0.31	0.35
Forward digit span	0.17	[0.05, 0.30]	.010		
Backward digit span	0.21	[0.08, 0.34]	.003		
Language [English]	0.90	[0.47, 1.34]	<.001	0.85	
Backward × Lang [English]	-0.37	[-0.66, -0.07]	.019		

Model 2 formula: imitation ~ melody + language + (language | participant) + (melody | item). Marginal $R^2 = 0.24$; Conditional $R^2 = 0.78$
 Model 3 formula: imitation ~ auditory pitch + language + (language | participant) + (1 | item). Marginal $R^2 = 0.23$; Conditional $R^2 = 0.76$
 Model 4 formula: imitation ~ forward + backward * language + (language | participant) + (1 | item). Marginal $R^2 = 0.26$; Conditional $R^2 = 0.77$

4. Discussion

The present study recruited L2 English speakers with no prior knowledge of Chinese to investigate whether individual differences in musical aptitude, auditory processing, and working memory affect their imitation of a familiar (English) and an unfamiliar (Chinese) L2. The results show that (a) overall, auditory processing and working memory significantly predict speech imitation skills; (b) music melody, auditory pitch processing abilities, and forward digit span are the significant components; and (c) the predictive value of overall auditory score and working memory score (especially, backward digit span) depends on whether the imitator was familiar with the target L2. In what follows, we discuss the results to address our research questions.

4.1 The predictive role of music aptitude, auditory processing, and working memory in L2 speech imitation

We first hypothesized that better cognitive abilities would imply better speech imitation skills. We found that only auditory processing abilities and working memory played a significant role, while music aptitude did not. This suggests that, globally, having a good ear and good short-term memory is more important than musical aptitude in imitating L2. Notably, the results do not rule out the predictive value of musical aptitude on L2 imitation skills but rather demonstrate that musical aptitude is less predictive if other cognitive factors are considered.

Regarding the role of language familiarity, we found that auditory processing only significantly predicted the imitation skills of the familiar L2, while working memory showed a stronger predictive value for the unfamiliar L2 than the familiar L2. This finding suggests that at the initial contact with an L2, a better short-term memory may weigh more than a better ear to capture and reproduce the phonetic details of nonnative sounds. However, when one has sufficient knowledge of the target L2, a good ear may help imitate fine-grained L2 phonetic details, reflected by a better imitation score. This finding adds new evidence to the field of L2 speech learning theory by demonstrating that short-term memory and auditory processing abilities are closely related to the degree of language familiarity.

4.2 The predictive role of specific components of music aptitude, auditory processing abilities, and working memory skills as well as the role of language familiarity

4.2.1 Musical aptitude

Within the domain of musical aptitude, melody perception ability appears to be the sole predictor of L2 imitation abilities, regardless of language familiarity. This finding is well in line with previous research that suggests that melodic perception abilities predict the imitation skills of unfamiliar languages (Delogu et al., 2010; Li et al., 2022, 2024) and extends the predictive value of melody to the imitation of familiar L2 as well. Interestingly, Li et al.'s (2022, 2024) studies showed that melodic perception abilities mainly predicted non-tone L1ers' imitation skills but not tone L1ers'. In our study, we only recruited participants from non-tone language backgrounds. More importantly, the coefficient of melodic perception abilities in our study ($\beta = 0.19$) was close to that of Li et al. (2024) ($\beta = 0.18$). Taken together, the two studies show that melodic perception abilities seem to be the most important musical aptitude in predicting non-tone L1er's L2 imitation skills.

By contrast, musical accent, pitch, and rhythm perception abilities were not significant predictors. Previous research has shown that musical accent may be more relevant for tone L1ers imitating unfamiliar L2, but not necessarily so for non-tone L1ers (Li et al., 2024). As we did not recruit any tone L1ers in the current study, musical accent ability was not a significant predictor. As for musical pitch and rhythm, these two components may be more related to specific domains, such as lexical tone (Bowles et al., 2016; Christiner et al., 2022; Li & Dekeyser, 2017), intonation (Yuan et al., 2019), word stress (Cason et al., 2020), speech rhythm (Boll-Avetisyan et al., 2017), and duration (Li et al., 2020), rather than speech imitation in general. Notably, Christiner and Reiterer's (2013) study showed that rhythmic perception ability predicted profes-

sional musicians' imitation skills in an unfamiliar language (Hindi). However, none of the participants in this study were musicians. Therefore, the findings of Christiner and Reiterer (2013) may be a confounding result of musical aptitude and musical expertise.

4.2.2 Auditory processing abilities

As proposed by the Auditory Precision Hypothesis-L2 (Saito, 2023), individuals with more precise auditory processing abilities would be better at producing L2 speech. This hypothesis pertains to actual L2 speech learning. Our data suggest that pitch processing ability accounts for L2 speech imitation skills, regardless of learners' prior knowledge of the target L2. This finding has thus contributed new evidence to Saito's hypothesis and suggests a potential differential role of the specific auditory domains that may dominate speech imitation skills in L2.

We did not find auditory processing abilities in duration or formant to be significant predictors of L2 imitation skills. It might well be that duration and formant discrimination abilities are domain-specific in affecting phonetic details of L2 speech production. For instance, duration discrimination ability can predict segmental or word stress accuracy but not intonation or fluency (Zheng et al., 2022). On the other hand, formant discrimination ability is more related to segmental learning (Zheng et al., 2022) but not necessarily to the imitation of a new language or accent. Therefore, these two factors may not affect overall speech imitation skills.

4.2.3 Working memory

According to the framework of working memory models (Baddeley, 2003; Baddeley et al., 1998, 2021), short-term memory is an essential component of the processing of verbal information, which supports L2 speech production. Our data showed that the forward digit span is a significant predictor of L2 speech imitation ability in both familiar and unfamiliar L2s. This supports the positive role of working memory in L2 speech production (Aliaga-García et al., 2011; Baker, 2008; Fortkamp, 2000; O'Brien et al., 2006, 2007; Trude & Tokowicz, 2011).

More importantly, specific working memory measures showed differential predictive values for familiar and unfamiliar L2s. The more challenging backward digit span predicted the imitation of an unfamiliar L2 better than a familiar L2. It might be that the backward digit span requires both information recall and information processing (Kormos & Sáfár, 2008), but the forward digit span mainly tests information recall. This distinction is important because imitating an unfamiliar L2 does not require semantic information processing, and imitators only need to recall a sequence of phonetic forms. However, as our participants can speak English, a better

information processing ability may allow them more cognitive space to process the speech signal and reproduce it in their “own” way, while neglecting the accent of the model speaker.

5. Conclusions and practical implications

The current study has theoretical and practical implications. First, we demonstrated that individual cognitive differences are key to predicting imitation skills in L2 speech production. It is relevant not only to newly exposed verbal materials but also to an already well-acquired L2. Among the cognitive measures, forward digit span, auditory pitch processing ability, and musical melody perception ability were positively correlated with L2 imitation skills in general. This finding suggests that in L2 teaching practice, training learners’ auditory processing abilities would lead to better learning outcomes (Saito, Petrova, et al., 2022) and making good use of the connections between music and speech (Cason et al., 2015; Liu & Liu, 2024) would contribute to L2 learning as well.

Second, our results contribute to theoretical models capturing the relationship between cognitive factors and speech production, such as the Multicomponent Working Memory model (Baddeley et al., 2021), the Auditory Precision Hypothesis-L2 (Saito, 2023), and the OPERA hypothesis (Patel, 2011). We expanded the scope of this field by adding a new factor, language familiarity. It seems that the specific components of the cognitive factors that play a role in speech imitation skills are subject to the typology of the imitator’s L1 (Li et al., 2024) and L2 under study (Christiner et al., 2018). Our study suggests that the imitator’s familiarity with the target L2 also makes a difference.

Third, in L2 teaching practice, attention should be paid to individual differences according to whether the learners are novice learners or experienced learners. For example, training auditory processing skills to improve pronunciation may be more effective for experienced learners than for beginners.

Finally, the current study has some limitations. To make the experiment at a reasonable length, we did not vary the typology of the L2 across familiarity. It would be interesting for a future study to employ a 2×2 design where both familiar and unfamiliar L2s feature two typologically different languages. Moreover, although efforts were made to include participants from a variety of L1 backgrounds, the moderate sample size and dual-only L2 conditions may limit generalizability to populations speaking other language combinations. Future studies may seek to test a greater number of diverse familiar and unfamiliar L2s.

To conclude, this study contributes to the body of L2 speech acquisition literature that emphasizes the significant role of individual cognitive differences—musical aptitude, auditory processing abilities, and working memory—in influencing speech imitation skills. More importantly, we demonstrate that the predictive effects of the cognitive measures depend on the imitators' familiarity with the target language, which provides insights into L2 teaching practice, as imitating an unfamiliar language well resembles the first encounter with a novel language. It is plausible that the relevant components of individual cognitive factors vary across the learners' experiences in the target L2 ranging from novice learners to experienced learners. Building on this, further research could tailor instruction by considering the cognitive factors in constructing teaching content, practice and feedback to learner's strengths. For instance, technology-mediated platforms could monitor cognitive load and auditory performance in real time which enables dynamic practice adjustments to optimize engagement and progress across diverse L2 learners. These teaching practices can lead to important implications regarding how individual differences and differentiated instruction contribute to L2 educational development.

Acknowledgements

This research is supported by the Basque Government through the BERC 2022-2025 program, by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation [CEX2020-001010/AEI/10.13039/501100011033], by the Spanish Ministry of Science, Innovation and Universities, the Spanish State Research Agency (MCIU/AEI/10.13039/501100011033), and the European Union "NextGenerationEU"/PRTR" through the Juan de la Cierva program [JDC2022-048729-I], and by the Norwegian Research Council through the Center of Excellence funding scheme [223265]. II, IM, PP, BR, and LX made equal contribution to this work. Their names are listed in alphabetic order of the last names. The authors thank Mr. Erick Chen for proofreading the Spanish version of the abstract.

References

Aliaga-García, C., & Mora, J. C. (2008). Perception and production of oral stops by Catalan/Spanish learners of English: A phonetic training experiment. In R. Monroy Casas & A. Sánchez Pérez (Eds.), *25 Años de Lingüística en España. Hitos y retos* (pp. 9–15). Universidad de Murcia. <http://dialnet.unirioja.es/servlet/articulo?codigo=4649737>

Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, 47(1), 1–14. <https://doi.org/10.2478/psicl-2011-0002>

Baddeley, A. (1992). Working Memory. *Science*, 255(5044), 556–559. <https://doi.org/10.1126/science.1736359>

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189–208. [https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158–173. <https://doi.org/10.1037/0033-295x.105.1.158>

Baddeley, A., Hitch, G., & Allen, R. (2021). A Multicomponent Model of Working Memory. In A. Baddeley, G. Hitch, & R. Allen, *Working Memory* (pp. 10–43). Oxford University Press. <https://doi.org/10.1093/oso/9780198842286.003.0002>

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27(5), 586–595. [https://doi.org/10.1016/0749-596X\(88\)90028-9](https://doi.org/10.1016/0749-596X(88)90028-9)

Baker, W. (2008). Social, experiential and psychological factors affecting L2 dialect acquisition. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 187–198). <https://scholarsarchive.byu.edu/facpub/5915>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear Mixed-Effects Models using {lme4}. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Berthier, M. L., & Lambon Ralph, M. A. (2014). Dissecting the function of networks underpinning language repetition. *Frontiers in Human Neuroscience*, 8, Article 727. <https://doi.org/10.3389/fnhum.2014.00727>

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, 930, 232–258. <https://doi.org/10.1111/j.1749-6632.2001.tb05736.x>

Boll-Avetisyan, N., Bhatara, A., & Höhle, B. (2017). Effects of musicality on the perception of rhythmic structure in speech. *Laboratory Phonology*, 8(1), 1–16. <https://doi.org/10.5334/labphon.91>

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, 66(4), 774–808. <https://doi.org/10.1111/lang.12159>

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica, 155*, 43–50. <https://doi.org/10.1016/j.actpsy.2014.12.002>

Cason, N., Marmursztejn, M., D'Imperio, M., & Schön, D. (2020). Rhythmic abilities correlate with L2 prosody imitation abilities in typologically different languages. *Language and Speech, 63*(1), 149–165. <https://doi.org/10.1177/0023830919826334>

Chen, N. F., Shivakumar, V., Harikumar, M., Ma, B., & Li, H. (2013). Large-scale characterization of mandarin pronunciation errors made by native speakers of European languages. In F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, & P. Perrier (Eds.), *Proceedings of INTERSPEECH 2013* (pp. 2370–2374). <https://doi.org/10.21437/Interspeech.2013-553>

Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System, 72*, 85–98. <https://doi.org/10.1016/j.system.2017.10.010>

Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology, 4*, 874. <https://doi.org/10.3389/fpsyg.2013.00874>

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience, 9*, 482. <https://doi.org/10.3389/fnhum.2015.00482>

Christiner, M., & Reiterer, S. M. (2018). Early influence of musical abilities and working memory on speech imitation abilities: Study with pre-school children. *Brain Sciences, 8*(9). <https://doi.org/10.3390/brainsci8090169>

Christiner, M., Rüdigger, S., & Reiterer, S. M. (2018). Sing Chinese and tap Tagalog? Predicting individual differences in musical and phonetic aptitude using language families differing by sound-typology. *International Journal of Multilingualism, 15*(4), 455–471. <https://doi.org/10.1080/14790718.2018.1424171>

Christiner, M., Serrallach, B. L., Benner, J., Bernhofs, V., Schneider, P., Renner, J., Sommer-Lolei, S., & Groß, C. (2022). Examining Individual Differences in Singing, Musical and Tone Language Ability in Adolescents and Young Adults with Dyslexia. *Brain Sciences, 12*(6), 744. <https://doi.org/10.3390/brainsci12060744>

Coumel, M., Christiner, M., & Reiterer, S. M. (2019). Second Language Accent Faking Ability Depends on Musical Abilities, Not on Working Memory. *Frontiers in Psychology, 10*, 257. <https://doi.org/10.3389/fpsyg.2019.00257>

Coumel, M., Groß, C., Sommer-Lolei, S., & Christiner, M. (2023). The Contribution of Music Abilities and Phonetic Aptitude to L2 Accent Faking Ability. *Languages*, 8(1), 68. <https://doi.org/10.3390/languages8010068>

Darcy, I., Mora, J. C., & Daidone, D. (2016). The Role of Inhibitory Control in Second Language Phonological Processing. *Language Learning*, 66(4), 741–773. <https://doi.org/10.1111/lang.12161>

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, 22(1), 46–61. <https://doi.org/10.1080/09541440802708136>

Delogu, F., & Zheng, Y. (2020). Beneficial Effects of Musicality on the Development of Productive Phonology Skills in Second Language Acquisition. *Frontiers in Neuroscience*, 14, 618. <https://doi.org/10.3389/fnins.2020.00618>

Denk, T. (2023). *Digit span test*. <https://tools.timodenk.com/digit-span-test>

Fortkamp, M. B. M. (2000). *Working Memory Capacity and L2 Speech Production: An Exploratory Study* [Universidade Federal de Santa Catarina]. <http://repositorio.ufsc.br/xmlui/handle/123456789/78287>

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement version 0.84.1 [software]*. <https://cran.r-project.org/package=irr>

Ghazi-Saidi, L., & Ansaldo, A. I. (2017). Second Language Word Learning through Repetition and Imitation: Functional Networks as a Function of Learning Phase and Language Distance. *Frontiers in Human Neuroscience*, 11, 463. <https://doi.org/10.3389/fnhum.2017.00463>

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>

Jekiel, M., & Malarski, K. (2021). Musical Hearing and Musical Experience in Second Language English Vowel Acquisition. *Journal of Speech, Language, and Hearing Research*, 64(5), 1666–1682. https://doi.org/10.1044/2021_JSL-HR-19-00253

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism*, 11(2), 261–271. <https://doi.org/10.1017/S1366728908003416>

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS ONE*, 7(12). <https://doi.org/10.1371/journal.pone.0052508>

Li, M., & Dekeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, 39(4), 593–620. <https://doi.org/10.1017/S0272263116000358>

Li, P., Baills, F., & Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts. *Studies in Second Language Acquisition*, 42(5), 1015–1039. <https://doi.org/10.1017/S0272263120000054>

Li, P., & Xi, X. (2024). The perception of Spanish lexical stress by proficient Mandarin learners of Spanish. In Y. Chen, A. Chen, & A. Arvaniti (Eds.), *Proceedings of the 12th International Conference on Speech Prosody* (pp. 354–358). <https://doi.org/10.21437/SpeechProsody.2024-72>

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 23(5), 938–944. <https://doi.org/10.1017/S1366728918001153>

Li, P., Zhang, Y., Baills, F., & Prieto, P. (2024). Musical perception skills predict speech imitation skills: Differences between speakers of tone and intonation languages. *Language and Cognition*, 16(3), 647–665. <https://doi.org/10.1017/lang-cog.2023.52>

Li, P., Zhang, Y., Fu, X., Baills, F., & Prieto, P. (2022). Melodic perception skills predict Catalan speakers' imitation abilities of unfamiliar languages. In S. Frota, M. Cruz, & M. Vigário (Eds.), *Proceedings of the 11th International Conference on Speech Prosody* (pp. 876–880). <https://doi.org/10.21437/SpeechProsody.2022-178>

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press. <https://doi.org/10.5070/L461005210>

Liu, X., & Liu, Y. (2024). Music Rhythmic Cueing for the Production of Non-native Speech Rhythm: Evidence from Chinese Learners of French. *Journal of Psycholinguistic Research*, 53(1), 1–23. <https://doi.org/10.1007/s10936-024-10044-1>

Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 101(2), 343–352. <https://doi.org/10.1037/0033-295X.101.2.343>

Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: A review. *Frontiers in Psychology*, 2, 321. <https://doi.org/10.3389/fpsyg.2011.00321>

Mizera, G. J. (2006). *Working memory and L2 oral fluency* [University of Pittsburgh]. <https://core.ac.uk/download/pdf/12207478.pdf>

Mora, J. C., & Darcy, I. (2023). Individual differences in attention control and the processing of phonological contrasts in a second language. *Phonetica*, *80*(3–4), 153–184. <https://doi.org/10.1515/phon-2022-0020>

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). *Tools for second language speech research and teaching* [Dataset]. <https://doi.org/10.17616/R31NJNAX>

Murljadic, M. (2020). *Musical ability and accent imitation* [University of Connecticut]. Honors Scholar Theses. https://opencommons.uconn.edu/srhonors_theses/688

O'Brien, I., Segalowitz, N., Collentine, J. O. E., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, *27*(3), 377–402. <https://doi.org/10.1017/S0142716406060322>

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(4), 557–581. <https://doi.org/10.1017/S027226310707043X>

Patel, A. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology*, *2*, 142. <https://doi.org/10.3389/fpsyg.2011.00142>

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions B*, *370*(1664), 1–8. <https://doi.org/10.1098/rstb.2014.0090>

Posedel, J., Emery, L., Souza, B., & Fountain, C. (2012). Pitch perception, working memory, and second-language phonological production. *Psychology of Music*, *40*(4), 508–517. <https://doi.org/10.1177/0305735611415145>

R Core Team. (2014). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <http://www.r-project.org/>

Reiterer, S. M., Hu, X., Sumathi, T. A., & Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology*, *4*, 782. <https://doi.org/10.3389/fpsyg.2013.00782>

Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing Auditory Precision Hypothesis-L2. *Language Teaching*, *56*(4), 522–538. <https://doi.org/10.1017/S0261444822000453>

Saito, K., Cui, H., Suzukida, Y., Dardon, D. E., Suzuki, Y., Jeong, H., Révész, A., Sugiura, M., & Tierney, A. (2022). Does domain-general auditory processing uniquely explain the outcomes of second language speech acquisition, even once cognitive and demographic variables are accounted for? *Bilingualism: Language and Cognition*, 25(5), 856–868. <https://doi.org/10.1017/S1366728922000153>

Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020). Domain-general auditory processing as an anchor of post-pubertal second language pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language*, 115, 104168. <https://doi.org/10.1016/j.jml.2020.104168>

Saito, K., Kachlicka, M., Suzukida, Y., Petrova, K., Lee, B. J., & Tierney, A. (2022). Auditory precision hypothesis-L2: Dimension-specific relationships between auditory processing and second language segmental learning. *Cognition*, 229, 105236. <https://doi.org/10.1016/j.cognition.2022.105236>

Saito, K., Macmillan, K., Kroeger, S., Magne, V., Takizawa, K., Kachlicka, M., & Tierney, A. (2022). Roles of domain-general auditory processing in spoken second-language vocabulary attainment in adulthood. *Applied Psycholinguistics*, 43(3), 581–606. <https://doi.org/10.1017/S0142716422000029>

Saito, K., Petrova, K., Suzukida, Y., Kachlicka, M., & Tierney, A. (2022). Training auditory processing promotes second language speech acquisition. *Journal of Experimental Psychology: Human Perception and Performance*, 48(12), 1410–1426. <https://doi.org/10.1037/xhp0001042>

Saito, K., Sun, H., & Tierney, A. (2020). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*, 41(5), 1083–1112. <https://doi.org/10.1017/S0142716420000491>

Saito, K., & Tierney, A. (2024). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, 46(4), 1206–1230. <https://doi.org/10.1017/S027226312200047X>

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17(8), 675–681. <https://doi.org/10.1111/j.1467-9280.2006.01765.x>

Sun, H., Saito, K., & Tierney, A. (2021). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, 43(3), 551–573. <https://doi.org/10.1017/S0272263120000649>

Trude, A. M., & Tokowicz, N. (2011). Negative Transfer From Spanish and English to Portuguese Pronunciation: The Roles of Inhibition and Working Memory. *Language Learning*, 61(1), 259–280. <https://doi.org/10.1111/j.1467-9922.2010.00611.x>

Voeten, C. C. (2021). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression* [Computer software]. <https://cran.r-project.org/package=buildmer>

Wiley, R. W., & Rapp, B. (2019). Statistical analysis in Small-N Designs: Using linear mixed-effects modeling for evaluating intervention effectiveness. *Aphasiology*, 33(1), 1–30. <https://doi.org/10.1080/02687038.2018.1454884>

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 41(1), 5–32. <https://doi.org/10.1017/S0272263117000316>

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, 1400(1), 33–45. <https://doi.org/10.1111/nyas.13410>

Zheng, C., Saito, K., & Tierney, A. (2022). Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude. *Second Language Research*, 38(3), 477–497. <https://doi.org/10.1177/0267658320978493>

Appendix A. Imitation Task Stimuli

English phrases	Chinese phrases
Short phrases	
They just washed the dirty sheets.	可爱猫咪 /kʰɿ³. ai⁴. mɑʊ¹. mi¹/ Lovely cat.
He put down knives, but no forks.	两只老虎 /lʲiɑŋ³. tʂʰ¹. laʊ². xu³/ Two tigers.
He booked a room for four nights.	纯色裙子 /tʂʰuən². sɿ⁴. tɛʰyn². tsɿ/ Solid color skirt.
Medium-length phrases	
I used to go cycling once a week.	那电影很好看 /na⁴. tʲen⁴. ɿŋ³. xən². xaʊ³. kʰan⁴/ That movie is great.
Are there umbrellas I can borrow?	他喜欢听音乐 /tʰa¹. ɛi³. xwan¹. tʰɿŋ¹. m¹. qœ⁴/ He likes listening to music.
She is the owner of the brewery.	我们需要知识 /uɔ³. mən. ɛy¹. jaʊ⁴. tʂʰ¹ ʂʰ/ We need knowledge.
Long phrases	
This is the thing I was thinking of buying.	度过了美好的一天 /tu⁴. kuɔ⁴. lɿ. mer². xaʊ³. tɿ. i⁴. tʰjɛn¹/ It's been a nice day.
They stole the crown jewels at the coronation?	郊外有一片白桦林 /tɛjaʊ¹. uar⁴. joʊ³. i². pʰjɛn⁴. par². xwa⁴. lɿn²/ There is a birch forest in the suburbs.
The sixth glass of beer made him very nauseous.	警察让出租车停下 /tɛɿŋ³. tʂʰa². zɑŋ⁴. tʂʰu¹. tsu¹. tʂʰɿ¹. tʰɿŋ². ɛja⁴/ The police stopped the taxi.

Note. Lexical tones are marked by numbers (1-4). The neutral tone is unmarked.

