


ARTICLE

Musical perception skills predict speech imitation skills: differences between speakers of tone and intonation languages

Peng Li^{1,2} , Yuan Zhang², Florence Bails^{2,3} and Pilar Prieto^{2,4}

¹Centre for Multilingualism in Society across the Lifespan (MultiLing), Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway; ²Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain; ³Institute for Linguistics- Phonetics, University of Cologne, Cologne, Germany; ⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Corresponding author: Peng Li; Email: peng.li@iln.uio.no

(Received 30 May 2023; Revised 02 October 2023; Accepted 04 October 2023)

Abstract

The ability to imitate speech is linked to individual cognitive abilities such as working memory and the auditory processing of music. However, little research has focused on the role of specific components of musical perception aptitude in relation to an individual's native language from a crosslinguistic perspective. This study explores the predictive role of four components of musical perception skills and working memory on phonetic language abilities for speakers of two typologically different languages, Catalan (an intonation language) and Chinese (a tone language). Sixty-one Catalan and 144 Chinese participants completed four subtests (accent, melody, pitch and rhythm) of the Profile of Music Perception Skills, a forward digit span task and a speech imitation task. The results showed that for both groups of participants, musical perception skills predicted speech imitation accuracy but working memory did not. Importantly, among the components of musical perception skills, accent was the only predictive factor for Chinese speakers, whereas melody was the only predictive factor for Catalan speakers. These findings suggest that speech imitation ability is predicted by musical perception skills rather than working memory and that the predictive role of specific musical components may depend on the phonological properties of the native language.

Keywords: intonation languages; musical aptitude; speech imitation; tone languages; working memory

Peng Li and Yuan Zhang made equal contributions to this work and share co-first authorship.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

1. Introduction

In the last few decades, there has been growing interest in the predictive role of individual differences in ‘phonetic language abilities’. Christiner and Reiterer (2018) defined ‘phonetic language abilities’ as ‘the capacity to imitate, mimic and pronounce spoken speech based on holistic judgments of human native speaker raters, judging imitated prosody as well as phonetic (segmental) aspects’ (p. 2). In particular, the differences between the cognitive function of working memory and music perception as individual aptitudes have received special attention. On the one hand, music and language are both related to human acoustic and sensory-motor systems and these common networks and processes have led to the hypothesis that music may influence language production (Patel, 2011). On the other hand, there is evidence that working memory capacity affects language processing and production (Christiner & Reiterer, 2018; Christiner et al., 2018, 2022). In what follows, we review the literature on the role of musicality and working memory in the perception and production of both familiar languages – including an L2 in the process of acquisition – and unfamiliar languages.

1.1. The role of musicality in the perception and production of familiar and unfamiliar languages

Although there is still debate on whether the mechanisms underlying speech and music perception are overlapping or rather dissociable, cognitive science has revealed a compelling and complex relationship between music and language. On the one hand, neuroscientific evidence revealed that the representations of music and non-music sounds are distinct in the auditory cortex (Boebinger et al., 2021; Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015, 2022). On the other hand, the processing of music and speech stimuli involves some overlapping brain areas (Patel, 2014; Peretz et al., 2015), because the response to music and speech activates a large overlapping portion of the auditory cortex (Rogalsky et al., 2011). Moreover, violating syntactic rules of speech and harmonic rules of music led to similar neural responses (Besson & Schön, 2001). Taken together, although the neuronal populations that respond to music and speech differ, they seem to occur in overlapping brain areas (Peretz et al., 2015). Therefore, it is reasonable to hypothesize that when neural networks are trained through extensive musical practice, they can help process acoustic information related to not only music but also speech with high precision (Patel, 2011, 2014).

An individual’s ability to perceive and produce the phonetic features of nonnative speech correlates with both their musical expertise and musical aptitude. Musical expertise is usually defined as the number of years of formal musical training (Zhang et al., 2020), and musical aptitude refers to the ability to intuitively learn, understand and appreciate music, a kind of inherent potential for learning music before formal music training (Law & Zentner, 2012). To capture this individual ability, researchers usually measure the participants’ ability to discriminate between differences in various components of music like rhythm and pitch. This is done by playing paired musical statements to the participants and then asking them to indicate whether the statements they heard were the same or different.

First, individuals with higher musical expertise tend to excel in processing and perceiving pitch in speech, for instance, in the discrimination and identification of L2 lexical tones (Cooper & Wang, 2012; Delogu et al., 2010; Gottfried et al., 2004; Lee &

Hung, 2008; Marie et al., 2011) or pitch deviations in L2 intonation (Marques et al., 2007; Martínez-Montes et al., 2013). In addition, experienced musicians show higher sensitivity to other aspects of speech processing such as rhythmic grouping (Boll-Avetisyan et al., 2016), speech stream segmentation (François et al., 2014), speech timing (Sadakata & Sekiyama, 2011), speech sound perception (Marie et al., 2011; Sadakata & Sekiyama, 2011) and even the perception of subsegmental features like voice onset time (Ott et al., 2011). More importantly, musicians outperform non-musicians in their ability to imitate unfamiliar languages as measured in terms of intelligibility (Delogu & Zheng, 2020), suprasegmental accuracy (Pei et al., 2016) and overall imitation accuracy (Murljacic, 2020; Pastuszek-Lipinska, 2008).

Second, regarding musical aptitude, musical perception skills play a predictive role in speech perception and production abilities. Slevc and Miyake (2006) showed that musical aptitude predicts receptive and productive L2 phonology, and their findings were confirmed and extended in subsequent studies. For instance, among nontonal language speakers, those who showed higher musical aptitude outperformed those who were less musically talented in identifying L2 lexical tones (Cooper & Wang, 2012). Learners with better musical aptitude also showed enhanced intelligibility in L2 speech imitation (Delogu & Zheng, 2020). Likewise, children with better musical aptitude were more likely to perceive changes in duration in both speech and music (Milovanov et al., 2009). Interestingly, the results of a study involving accent-faking tasks in which participants were asked to speak in their L1 while imitating an L2 accent suggested that participants with greater musical aptitude were also able to do this more easily, suggesting a correlation with overall phonological awareness (Coumel et al., 2019).

Unsurprisingly, musical production skills are associated with productive language abilities. For instance, adults and children with higher singing aptitudes performed better in imitating a series of unintelligible and unfamiliar speech sounds (Christiner & Reiterer, 2013, 2018) and better overall L2 pronunciation (Milovanov et al., 2008). A higher musical production aptitude may also have a positive effect on not only the production but also the perception of L2 speech. For instance, Li and Dekeyser (2017) and Slevc and Miyake (2006) measured participants' ability to produce tones by asking them to orally repeat the musical stimuli played to them using the syllable sequence 'lalala', their output being recorded and later rated by professional singers. Both studies found positive correlations between musical production aptitude and L2 speech production and perception skills.

1.2. Correlations between the specific components of musical aptitude and language perception and production

Musical aptitude is a multi-dimensional construct that consists of many components. Several studies have investigated how its separate components correlate to specific aspects of speech perception and production. Thus far, it is the rhythmic and pitch perception and production components that have received the most attention.

Rhythmic perception skills correlated with a more accurate perception of speech rhythm (Boll-Avetisyan et al., 2017), the production of L2 long and short vowels (Li et al., 2020), and the ability to imitate unfamiliar languages accurately (Christiner & Reiterer, 2013). By the same token, individuals with better rhythmic production skills (i.e., they are better able to reproduce musical rhythmic sequences) could

produce word stress more accurately and thus exhibited greater fluency in an L2 (Zheng et al., 2022), and also reproduced unfamiliar prosody more accurately, specifically in terms of stress-accent placement (Cason et al., 2020).

Musical pitch perception skills were associated not only with the perception of L2 lexical tones (Li & Dekeyser, 2017) and the production of lexical tones in unfamiliar languages (Christiner et al., 2022) but also with L2 pronunciation in general (Posedel et al., 2012). Importantly, pitch perception abilities may predict successful learning of L2 words with lexical tones (Bowles et al., 2016) and the production of L2 intonation (Yuan et al., 2019). However, when the learning target is not related to pitch, pitch perception skills do not correlate significantly with the ability to learn other phonetic features of an L2, such as vowel length (Li et al., 2020).

There is contradictory evidence regarding the role of other components of musical aptitude in phonetic language abilities. For example, melodic perception skills correlated with the perception of L2 lexical tones (Delogu et al., 2010) and the production of L2 intonation (Jekiel & Malarski, 2023; Yuan et al., 2019). However, a recent study found no significant correlation between melodic production skills and L2 speech production (Zheng et al., 2022). Similarly, accent and melody perception skills, but not tempo or tuning skills, significantly predicted the imitation performance of English regional variants by native English speakers (Murljacic, 2020). A recent study, however, has shown that while musical rhythm and pitch perception abilities alone could not predict accent-faking accuracy, singing abilities could (Coumel et al., 2023).

At the segmental level, different components of musical aptitude may correlate with the ability to produce challenging L2 sounds, although studies have yielded inconsistent results. For instance, the musical timing perception skills of Japanese students predicted their ability to imitate English /r-l/ contrasts accurately, whereas pitch, loudness and rhythmic perception skills did not (Dolman & Spring, 2014). While having good rhythmic perception skills positively correlated with the ability to produce challenging L2 vowels, this was not the case with melodic and pitch perception skills (Jekiel & Malarski, 2021).

Nevertheless, only a handful of studies have looked for correlations between the different components of musical aptitude and language phonetic abilities as reflected through individuals' abilities to imitate unknown languages, and these studies have yielded mixed results. The first study compared the predictive role of rhythm and pitch perception abilities in the production of familiar (English) and unfamiliar (Hindi) languages by German speakers and found that rhythmic – but not pitch – perception abilities significantly predicted the ability to imitate Hindi (Christiner & Reiterer, 2013). Later, Christiner et al. (2018) showed that the predictive role of specific music components on the imitation abilities of unfamiliar languages may be dependent on the typology of the target language. Specifically, they found that pitch perception ability predicted the imitation abilities in a tone language (Chinese) while rhythm perception ability predicted the imitation abilities in a stress language (Tagalog). Finally, pitch perception abilities could predict the imitation accuracy of Chinese tones by German speakers who had no prior knowledge of Chinese (Christiner et al., 2022).

To the best of our knowledge, no previous studies have assessed whether the native language of the participants can also modulate the predictive role of the musical aptitude components, as most of the studies included participants from a homogeneous L1 background. Christiner et al. (2018) tested participants with different native

languages including Bosnian, Serbian, Turkish and Macedonian. However, they did not test whether the participants' L1 influenced the predictive value of the different musical components. In other words, it remains an open question whether the results can be applied to speakers of different language typologies. Since tone languages manipulate pitch more on the lexical level than on the intonational level, tone language speakers might differ from intonation language speakers in their sensitivity to certain musical components. In fact, Chinese speakers have been shown to have finely tuned pitch perception skills similar to those of musicians (Bidelman et al., 2013).

Taken as a whole, this body of research suggests that it would be of interest to explore how speakers of tonal languages differ from intonation language speakers in terms of how their musical aptitude skills might be transferable to their processing and production abilities of L2s or unfamiliar languages.

1.3. The role of working memory in the perception and production of familiar and unfamiliar languages

Working memory refers to the temporary storage and simultaneous manipulation of information during cognitive processes, providing interfaces between perception, long-term memory and action. It is critical for higher cognitive functions such as planning, problem-solving and reasoning, as well as for processing and decoding speech and music (Schulze & Koelsch, 2012). In the context of L2 learning, working memory positively correlates with overall language proficiency (Kormos & Sáfár, 2008), vocabulary learning (Cheung, 1996) and grammar accuracy (Abdallah, 2010; O'Brien et al., 2006).

Regarding L2 speech learning, empirical research has not yet yielded consistent results on the predictive role of working memory. On the positive end, working memory related to the development of speech fluency (O'Brien et al., 2007), narrative skills (O'Brien et al., 2006) and overall speech proficiency as measured by complexity, accuracy and fluency (Fortkamp, 2000; Trude & Tokowicz, 2011). Working memory may also affect outcomes of L2 pronunciation training such as accuracy in the imitation of an English dialect (Baker, 2008) and the perceptual learning of individual vowels (Aliaga-Garcia et al., 2010). By contrast, some studies did not show significant correlations between working memory and aspects of L2 speech production, such as fluency (Mizera, 2006), overall pronunciation accuracy (Posedel et al., 2012, p. 201), intelligibility and accentedness (Slevc & Miyake, 2006) and the production of specific L2 features like duration (Li et al., 2020).

Likewise, mixed results were obtained on the role of working memory in predicting an individual's phonetic language abilities as manifested in their skill at imitating unfamiliar languages. Focusing first on the positive findings, working memory capacities have been shown to predict the imitation abilities of unfamiliar languages in both children (Christiner & Reiterer, 2018; Christiner et al., 2018) and adults (Christiner & Reiterer, 2013, 2018). By contrast, some recent studies have shown that while musical aptitude and singing abilities were significant predictors of phonological awareness as measured by an L2 accent-faking task (Coumel et al., 2019, 2023), working memory was not (Coumel et al., 2019). Also, working memory was not a significant predictor of the imitation of unfamiliar languages (Li et al., 2022). These results suggest that working memory capacity is a potential predictor of

individual differences in phonological awareness, although it might be less predictive than musical aptitude. Given the inconclusive findings in previous research, more evidence is needed to assess the predictive value of working memory. Therefore, it seems to be relevant to involve working memory in the investigation of phonetic language ability.

1.4. Goals of the present study

Considering the previous literature, further evidence is needed to determine which components of musical aptitude are better predictors of speech imitation abilities, and how they compare with working memory in this regard. Of those components, although the literature reviewed in [Section 1.1](#) identified rhythm and pitch as the most relevant components of musical aptitude in predicting phonetic language abilities, some results pointed to the relevance of melody and accent as well. Therefore, the present study will focus on those four components, namely, accent, melody, pitch and rhythm. In this study, then, we aim to investigate the predictive role of specific perceptive components of musical aptitude and working memory capacity on the speech imitation skills of two groups of participants with typologically different native languages, namely, Catalan (an intonation language) and Chinese (a tone language).

The present study poses the following two research questions:

- RQ1: Do musical perception skills predict phonetic language abilities better than working memory?
- RQ2: Which components of musical perception skills predict phonetic language abilities? Does the predictive effect of these components hold across speakers of typologically different languages?

For RQ1, we hypothesized that musical perception skills would be more predictive than working memory. Regarding RQ2, however, it is largely exploratory based on the typological differences between Chinese and Catalan. Chinese speakers showed pitch discrimination abilities similar to those of musicians (Bidelman et al., 2013), and Catalan speakers were sensitive to changes in specific parts of the pitch contour such as pitch accents and boundary tones (Prieto et al., 2015). Therefore, it would be reasonable to hypothesize that if speakers demonstrate musician-like expertise in one specific domain due to the prosodic properties of their L1, this musical skill will be less relevant to the imitation skills of unfamiliar languages compared to other components.

2. Methods

2.1. Participants

We recruited 144 Chinese-speaking middle-school students (80 females, mean age 13.93 years) from China and 61 Catalan-speaking undergraduate students (54 females, mean age 19.70 years) from Spain. All the participants reported having normal hearing and no speech impairment and had no prior exposure to the languages that were used in the speech imitation task. No participant had received musical training in voice, or a musical instrument was trained for more than half a

year. Thus, all the participants were considered to have essentially no musical expertise. The participants and their legal guardians, in the case of a minor, gave prior written consent allowing speech data collected from them to be used for academic purposes.

2.2. Materials

The experiment consisted of three tasks: a battery of tests assessing musical perception skills consisting of subtests for accent, melody, pitch and rhythm; a forward digit span task to measure working memory; and a speech imitation task with sentences in six languages that were unfamiliar to the participants to assess speech imitation skills.

2.2.1. Musical perception skills tests

To measure musical perception skills, we opted for the Profile of Music Perception Skills (PROMS; Law & Zentner, 2012), which is free online and provides an objective assessment of musical perception skills in various components such as pitch, rhythm, melody, accent, timbre, tempo and harmony. PROMS can be tailored to specific research needs in terms of both skill components and duration of the task (i.e., there are micro, mini, short and full versions), with even the short version producing reliable test scores and good internal consistency (Zentner & Strauss, 2017).

For the present study, we chose the short versions of the PROMS subtests measuring accent, melody, pitch and rhythm. The accent subtest measured the participants' ability to detect emphasis in rhythmic patterns with isometric notes varying in intensity. The melody subtest included monophonic rhythms. The pitch subtest used pure tones and varied pitch differences. The rhythm subtest had two-bar notes with constant intensity but varying duration. In all the subtests, participants were asked to detect differences between paired auditory stimuli, where the differences ranged from obvious to subtle.

2.2.2. Forward digit span task

Digit span is a measure of working memory, which belongs to the cognitive system that allows for the temporary storage of information (Baddeley, 2003). In order to keep the experiment a reasonable length, we selected a forward digit task, meaning that participants were only asked to repeat a sequence of digits in the order in which they had appeared and were not expected to try to repeat them in reverse order (a cognitively more challenging task). Adapting Woods et al.'s (2011) method, we used WinSCP software to develop an online test. The test was based on a script written by Eisenberg et al. (2017) and modified by Navarro Pérez and Rohrer (2020).

2.2.3. Speech imitation task

A total of six languages belonging to different language typologies were selected for the speech imitation task, with two sentences taken from each language. For L1 Chinese participants, the six target languages were Catalan, Hebrew, Japanese, Tagalog, Turkish and Vietnamese, whereas for L1 Catalan participants, we replaced Catalan with Chinese. The syllable count of the sentences varied from six to twelve. Table 1 lists all the sentences with English translations. It is important to point out

Table 1. Target sentences of the six languages and English gloss for these sentences used in the speech imitation task

Target sentences	English gloss
Catalan Els Jocs Olímpics d'hivern de Pyeongchang. Avui fa un dia molt bonic.	Pyeongchang Winter Olympic Games. It's a nice day today.
Chinese 今天是个好天气。 平昌冬季奥运会。	The weather is nice today. Pyeongchang Winter Olympics.
Hebrew שלום. שמי אלון ואני סטודנט. היום הוא יום יפה, והשמש זורחת.	Hello. My name is Alon and I am a student. Today is a beautiful day, and the sun is shining.
Japanese 会社にいらっしゃいますか。 食事していません。	Are you in the company? I haven't eaten.
Turkish Ali hayır dedi. Özge ona çarpılmıştı.	Ali said no. Özge had been lovestruck by him.
Russian мы работаем в офисе. эта газета лежит на столе.	We are working in the office. This newspaper is on the table.
Vietnamese Rất vui được gặp bạn. Làm ơn cho tôi mượn tờ giấy.	Nice to see you. Please lend me a piece of paper.

that the goal of the speech imitation test was to obtain an overall score of speech imitation abilities based on widely diverse phonetic targets; it was not designed to assess the participants' ability to imitate a specific language.

Seven native speakers (one for each language) were audio-recorded in a sound-proof room as they read each of the two sentences four times in a row. Afterward, the clearest tokens of the four recordings were selected as the target stimuli. The audio recordings were edited with Audacity and uploaded onto the Alchemer online survey platform (www.alchemer.com), where they constituted the auditory stimuli that participants would first hear and then repeat.

2.3. Procedure

After signing the written consent form, each participant carried out the full sequence of tasks, namely, musical skills subtests, forward digit span task, speech imitation task, online on a laptop, working individually and in a silent room. The full procedure lasted around 30 min per participant.

2.3.1. PROMS-S test battery

First, the participants did the PROMS-S subtests for accent, melody, pitch and rhythm, with each subtest containing eight to ten trials of varying degrees of difficulty. In each trial, participants first listened twice to the same stimulus (the 'referent'). After a short interval, they listened to a comparison stimulus (the 'comparison'). Participants were required to indicate whether the comparison stimulus differed from the referent stimulus or not and choose one answer from five

options: *definitely different, probably different, I don't know, probably the same and definitely the same*. The PROMS-S test battery lasted approximately 20 min.

2.3.2. Forward digit span task

Participants were then shown a link on the laptop screen to access the STM test. The task consisted of 14 trials. For each trial, participants were first presented with a sequence of digits appearing consecutively in the center of the screen and were then asked to replicate the sequence they had seen using the laptop keyboard, pressing the 'Enter' key when finished to proceed to the next trial. The number of digits in each sequence differed, with the first trial showing a sequence of only three digits. If the participants were able to replicate the three-digit sequence successfully, the program showed them a four-digit sequence, then a five-digit sequence, and so on. If the participant failed to correctly replicate a sequence of two trials in a row, the program reduced the length of the sequence by one digit. The task ended with the fourteenth trial regardless of how many digits had been presented in the last trial. The system automatically calculated and recorded participants' scores. The full task lasted approximately 5 min.

2.3.3. Speech imitation task

Finally, still working with the laptop, the participants proceeded to the online testing platform Alchemer to complete the speech imitation task. This involved listening to each model sentence twice and then imitating each sentence once. The 12 stimulus sentences (2 tokens \times 6 languages) were presented to the participants randomly and no translations were provided. The speech imitation test lasted approximately 5 min. Participants' oral output was recorded through a professional-quality audio recorder placed in front of them and activated by the experimenters at the outset of the speech imitation task. In total, 2,460 recordings were obtained of sentences being imitated [(144 Chinese participants + 61 Catalan participants) \times 6 languages \times 2 sentences].

2.4. Data coding

From the PROMS-S test battery results, a composite musical perception score was calculated by aggregating the scores of the four subsets (accent, melody, pitch and rhythm), which were automatically generated by the PROMS platform according to the following criteria. Whenever the participant correctly identified a comparison stimulus as being 'definitely' the same as or different from the referent stimulus, they were awarded two points; if they correctly identified the comparison stimulus as 'probably' the same or different, they were awarded one point. A wrong answer or 'I don't know' received 0 points. The score for each subtest was the sum of the scores for all items.

As noted above, scores on the forward digit span task were generated automatically by the program in WinSCP following the guidelines by Woods et al. (2011).

A score for participants' speech imitation ability was obtained as follows. First, the recordings of the participant imitating the two prompt sentences in each language were assessed perceptually by three native speakers of that language (7 languages \times 3 raters). Each rater judged how closely the participant approached native-like pronunciation on a 9-point Likert scale, with '1' indicating completely non-native or

Table 2. Inter-rater reliability as measured by intraclass correlation coefficients [lower bound, upper bound] for each of the languages imitated, broken down by participant group

	Catalan speakers	Chinese speakers
Catalan	–	0.86 [0.83, 0.89]
Chinese	0.87 [0.82, 0.90]	–
Hebrew	0.92 [0.90, 0.94]	0.72 [0.65, 0.77]
Japanese	0.88 [0.84, 0.91]	0.71 [0.65, 0.77]
Turkish	0.89 [0.85, 0.92]	0.86 [0.83, 0.89]
Russian	0.72 [0.62, 0.80]	0.91 [0.88, 0.92]
Vietnamese	0.93 [0.90, 0.95]	0.67 [0.60, 0.74]

unintelligible pronunciation and ‘9’ fully native-like pronunciation. Before performing the evaluations, all raters underwent a brief training session to try to ensure some consistency in the criteria they applied when rating. They were first given instructions about how to rate, it being emphasized that they were to rate recordings based on their overall impression of the speaker’s pronunciation rather than by focusing on elements such as specific phonemes. Raters were also instructed that a rating of 1 (the minimum) should be assigned to recordings where participants had produced only a small number of syllables because this constituted insufficient information to form a valid opinion. Raters then practiced by evaluating six sample recordings that were not part of the current experiment. The resulting ratings were compared, and whenever a sharp discrepancy among ratings was detected, this was discussed among the raters until a consensus was reached on the most appropriate rating. The same training procedure was carried out for each of the seven groups of language raters.

The raters then proceeded to rate the recordings of participants, working independently and in isolation, their ratings being recorded directly on the Alchemer online platform. This task required on average 90 min. After the ratings were completed, inter-rater reliability (intraclass correlation coefficients, ICCs) between the three raters of each language was checked using the *icc()* function from the *irr* package, version 0.84.1 (Gamer et al., 2019) in the R program, version 4.2.2 (R Core Team, 2014). The ICC was obtained from a series of mean ratings ($k = 3$), consistency, and two-way mixed-effects models. Most of the results (Table 2) showed an acceptable ($ICC > 0.7$) to excellent ($ICC > 0.9$) estimated mean ICC across the three raters for each of the six languages imitated by the two groups of participants (see Koo & Li, 2016 for the interpretation of ICC). Only the Vietnamese items imitated by Chinese speakers showed an estimated ICC below the 0.7 threshold due to the exclusion of the items with an imitation score of 1. If calculated without data exclusion (see Section 2.5), the mean ICC for Vietnamese imitated by Chinese speakers was 0.75 [0.70, 0.80]. We thus concluded that the shortfall of data here would not affect the overall validity of our analysis. Finally, we averaged the ratings of the three raters for each item to create a mean speech imitation score (henceforth ‘imitation score’) for the follow-up analysis.

2.5. Statistical analyses

Four linear mixed models (LMM) were built to analyze the predictive role of musical perception abilities and working memory using the *lmer()* function from the *lme4* package, version 1.1.31 (Bates et al., 2015) in R. Models 1 and 2 addressed RQ1 for Catalan speakers and Chinese speakers, respectively. Similarly, models 3 and

4 addressed RQ2 separately for each participant group. In all four models, the dependent variable was the speech imitation score. In models 1 and 2, the independent variables were the composite musical perception score and working memory score; whereas in models 3 and 4, the independent variables were the subtest scores for accent, melody, pitch and rhythm separately, and working memory score. Scores for all variables were transformed into *z*-scores. Specifically for the speech imitation data, before *z*-score transformation, all items that had obtained a mean rating of 1 (e.g., the three raters gave the score 1, meaning that the recording offered a too small speech sample to assess) were excluded from further analysis. In this way, 2 out of 732 speech recordings (0.3%) by Catalan participants and 121 out of 1,725 speech recordings (7%) by Chinese participants were excluded.

To select the best-fitting models, we built four full models including all the possible random slopes for the two random intercepts: participant and item. Here, item refers to the 12 sentences regardless of the language. We chose not to treat specific language as a fixed or random effect for several reasons. First, we were interested in participants' overall ability to imitate unfamiliar languages and not whether they could imitate one language better than another. Second, for each of the six target languages, participants were asked to imitate only two short sentences. As we were not interested in the by-language variance, we decided to treat each sentence as a single item when building the statistical models.

We then ranked all the possible models from the full model to the null model using the *buildmer()* function from the *buildmer* package, version 2.8 (Voeten, 2021). The best-fitting models were the best-ranking models without singular fit issues. As a result, model 1 (Catalan speakers) involved a random intercept of item with a random slope of working memory and a random intercept of participant with a random slope of musical perception score. Model 2 (Chinese speakers) involved a random intercept of item with a random slope of working memory and a random intercept of participant. Model 3 (Catalan speakers) involved a random intercept of participant with random slopes for working memory score and rhythm score, and a random intercept of item with a random slope for working memory score. Model 4 (Chinese speakers) involved a random intercept of item with random slopes for working memory score and pitch score, and a random intercept of participant.

3. Results

Table 3 summarizes the descriptive data for all the variables on their original scale from the Catalan and Chinese participants.

3.1. RQ1: Do musical perception skills predict phonetic language abilities better than working memory capacity?

The results of models 1 and 2 (Table 4) revealed a significant main effect of musical perception score (both $p < 0.05$), which means that participants' musical perception abilities significantly predicted their speech imitation abilities, for both Catalan and Chinese speakers. As for the role of working memory, there was no significant main effect in either model. This suggests that working memory is not a significant predictor of speech imitation abilities for either Catalan or Chinese speakers.

Table 3. Means (*M*), standard deviations (*SD*) and range of the scores of musical perception skills, accent, melody, pitch, rhythm, working memory and speech imitation

	Catalan speakers		Chinese speakers	
	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range
Musical perception score ^a	17.57 (4.79)	10.5–31	15.27 (4.14)	2.5–26
Accent	4.75 (1.65)	2–9.5	4.2 (1.78)	0–10
Melody	4.8 (1.73)	1.5–8.5	4.1 (1.49)	0–9.5
Pitch	3.12 (1.28)	1–6.5	3.6 (1.66)	0–8.5
Rhythm	4.89 (1.46)	1.5–7.5	3.38 (1.25)	0–7
Working memory score	7.01 (1.36)	4.77–11.17	7.99 (1.93)	2.5–16.5
Speech imitation score	5 (1.77)	1.33–9	2.8 (1.25)	1.33–7.67

^aMusical perception score is the sum of accent, melody, pitch, and rhythm scores.

Table 4. Results from the LMMs predicting imitation score with musical perception score, and working memory score as fixed effects and participant and item as random effects, broken down by participant group

Predictor	Fixed effects				Random effects	
	β	<i>SE</i>	<i>t</i>	<i>p</i>	By participant <i>SD</i>	By item <i>SD</i>
Model 1 (Catalan speakers)						
(Intercept)	0.00	0.16	0.01	0.996	0.13	0.29
Musical perception	0.19	0.06	3.20	0.001	0.01	–
Working memory	0.08	0.06	1.18	0.240	–	0.01
Model 2 (Chinese speakers)						
(Intercept)	–0.03	0.18	–0.17	0.867	0.10	0.37
Musical perception	0.08	0.03	2.63	0.009	–	–
Working memory	0.08	0.04	1.72	0.086	–	0.01

Note: Estimates (β) represent the change in speech imitation score resulting from a change in each fixed factor. Significant results are bolded.

3.2. RQ2: Which components of musical perception skills predict phonetic language abilities, and does the predictive effect of these components hold across speakers of typologically different languages?

As for the predictive role of the specific components of musical perception skills, model 3 (Table 5) and model 4 (Table 6) revealed different results. Model 3 showed that melody was the only significant predictor of Catalan speakers' imitation ability, whereas model 4 revealed that accent was the only significant predictor of Chinese speakers' imitation ability (both $p < 0.05$).

4. Discussion and conclusions

The present study examined (RQ1) the role of two cognitive individual factors, namely, musical perception skills and working memory capacity, in predicting phonetic language abilities; and (RQ2) whether the predictive effect of specific components of musical perception skills is subject to the speakers' native languages. The typologically different languages included Catalan (an intonation language) and Chinese (a tone language).

Table 5. Catalan-speaking participants: Results from the LMMs predicting imitation score with musical perception score and working memory score as fixed effects and participant and item as random effects

Predictor	Fixed effects				Random effects	
	β	<i>SE</i>	<i>t</i>	<i>p</i>	By participant	By item
					<i>SD</i>	<i>SD</i>
(Intercept)	0.00	0.16	0.01	0.995	0.11	0.29
Accent	0.06	0.07	0.88	0.379	–	–
Melody	0.18	0.06	2.81	0.005	–	–
Pitch	–0.06	0.06	–0.91	0.361	–	–
Rhythm	0.05	0.07	0.66	0.512	0.02	–
Working memory	0.04	0.06	0.61	0.545	0.02	0.01

Note: Estimates (β) represent the change in speech imitation score resulting from a change in each fixed factor. Significant results are bolded.

Table 6. Chinese-speaking participants: Results from the LMMs predicting imitation score with musical perception score and working memory score as fixed effects and participant and item as random effects

Predictor	Fixed effects				Random effects	
	β	<i>SE</i>	<i>t</i>	<i>p</i>	By participant	By item
					<i>SD</i>	<i>SD</i>
(Intercept)	–0.03	0.18	–0.17	0.866	0.10	0.37
Accent	0.07	0.03	2.17	0.030	–	–
Melody	–0.00	0.04	–0.06	0.953	–	–
Pitch	0.04	0.04	0.91	0.362	–	0.01
Rhythm	0.02	0.03	0.47	0.636	–	–
Working memory	0.08	0.05	1.76	0.078	–	0.01

Note: Estimates (β) represent the change in speech imitation score resulting from a change in each fixed factor. Significant results are bolded.

Regarding RQ1, we found that general musical perception skills – but not working memory capacity – predicted the imitation abilities of unfamiliar languages in the two groups of speakers. This is in line with the results of previous research showing that musical perception skills correlated with phonetic language abilities. In this regard, our findings add further cross-linguistic evidence that the phonetic language abilities of speakers of both intonation languages, like Catalan, and tone languages, like Chinese, are moderately predicted by their general musical aptitude, supporting the hypothesis that there is cognitive overlap between music and language (Chobert & Besson, 2013; Milovanov & Tervaniemi, 2011; Peretz et al., 2015).

We did not find working memory to significantly predict speech imitation abilities for either participant group. This is in line with previous research showing the limited utility of working memory capacity for predicting phonetic language abilities (Coumel et al., 2019; Li et al., 2022). Our null results regarding working memory capacity do not match the results of several comparable studies that found working memory to be a significant predictor of the ability to imitate unfamiliar languages (Christiner & Reiterer, 2018; Christiner et al., 2018, 2022). In our view, there are two possible explanations for this inconsistency. First, while participants in some of the previous work that highlighted the importance of working memory were young children (e.g., 5-year-olds in Christiner & Reiterer, 2018; 9-to-10-year-olds in

Christiner et al., 2018), our participants were adolescents and young adults. The role played by the working memory variable might conceivably be more evident in younger children than in older individuals. Second, our target sentences in the speech imitation task were not long and did not vary a great deal in length, with a mean syllable count of 8.5. The mean syllable count was close to the working memory scores of both groups of participants (Catalan speakers: 7.01 and Chinese speakers: 7.99). This means that working memory may not play a significant role when the target sentence length in the speech imitation task is similar to the participants' working memory span, as the demands of the imitation task do not exceed the participants' working memory capacity. Future research may want to control for the phonological length factor and adapt the length of target stimuli to exceed the working memory capacities of participants.

With respect to RQ2, our results contributed cross-linguistic data on which specific components of music perception abilities were predictors of phonetic language aptitude. Interestingly, the significant predictors of the two groups of participants were not the same. Specifically, the only predictive musical component of phonetic language aptitude for Chinese speakers was musical accent perception, while that for Catalan speakers was melody perception skills. In our view, this contrast can be explained by the differing prosodic nature of these two languages, Chinese being a tonal language and Catalan being an intonation language. On the one hand, since Chinese speakers already showed excellent pitch perception skills, which can be equated to those of musicians (Bidelman et al., 2013), we expected that other music perception skills might be more discriminatory in this population. It thus makes sense that the accent component was more discriminatory in this population since the strong–weak prominence contrast assessed by the PROMS accent subtest in Chinese is less phonologically relevant for Chinese speakers (Duanmu, 2007) than for stress language speakers like Catalan (Wheeler, 2005), where stress is an important feature of the phonology. Therefore, that Chinese participants, who were better at detecting strong–weak contrasts in music (i.e., the accent component), would be more sensitive to strong–weak contrasts in the imitation of unfamiliar speech as well and thus reproduced prominence differences better in speech. On the other hand, for Catalan participants, we would expect that musical components like accent differences, which are phonologically relevant in this intonation language would be less discriminatory in predicting speech imitation abilities. This was borne out by our results, where Catalan participants, who discriminated better across melodies of different musical phrases, as shown by the melody component in PROMS, were better at imitating unfamiliar speech. This ability may not have been as crucial for Chinese speakers, who are already trained to detect subtle melodic and pitch changes in their language. Though Catalan is an intonation-based language, it is not sentence-level pitch changes that are discriminatory but rather smaller-scale pitch accentual contrasts (Prieto et al., 2015). Catalan speakers are thus not experts in detecting fine-grained intonational differences at the sentence level; rather, their phonological expertise lies in detecting changes in pitch, duration and intensity in very specific parts of the contour (i.e., pitch accents and boundary tones). These results imply a skill transfer from the specific prosodic patterns of the L1 to the ability to detect those contrasts in musical phrases. Prosodic phonological abilities that are not specifically trained in the L1 are the most predictive of speech imitation abilities.

Following up on these findings, our results add new evidence to previous studies on the specific role of language background and musical aptitude skills in the

prediction of phonetic language abilities. In our study, melody and accent appear as the significant predictors. Yet since very few previous studies have included the perception abilities specifically related to accent and melody as components in their musical skills tests, we cannot make direct comparisons with other research. The small number of studies that have considered these components have focused on speech perception (Delogu et al., 2010), L2 intonation training (Yuan et al., 2019), and the production of challenging L2 sounds (Dolman & Spring, 2014). The only study involved cross-linguistic design was Christiner et al. (2018), which showed the predictive role of specific music components is dependent on the typology of the target languages being imitated, but the cross-linguistic design did not vary the speaker's L1 backgrounds. Our study thus provided new cross-linguistic evidence that suggests that speakers of different L1 backgrounds may be positioned differently with respect to the role of the various musical aptitude components in phonetic language abilities.

The present study suffers from several limitations. First, our measures of musical aptitude were based on perceptive abilities only. In the future, it would be of interest to use measures of productive abilities to look for links between music and language by comparing, for example, singing skills with speech across language typologies. Second, phonetic language abilities in our study were assessed in terms of participants' ability to imitate languages with which they were unfamiliar. It would be worthwhile to replicate the current study contrasting unfamiliar and familiar languages, or an L2 that the participants are learning. Doing so might yield results that would be of considerable utility to the field of second language acquisition. Finally, it is worth noting that due to human resource limits, we recruited more Chinese speakers ($N = 144$) than Catalan speakers ($N = 61$) and the two groups of participants differed in age as well (Chinese = 13.93 and Catalan = 19.7). Although both groups are young individuals, the differences in age and number of participants may have a potential influence on the results. Especially, adolescence is a crucial age for cognitive development (Müllensiefen et al., 2022). Future studies may want to replicate the current study with more comparable groups of participants in sample size, age and gender.

To conclude, the results of the present study constitute new cross-linguistic evidence that music and speech share common processes in the brain. More specifically, our findings show that the ability of specific components of musical perceptive aptitude to predict an individual's ability to imitate unfamiliar languages may be modulated by the prosodic specificities of the individual's native language, a finding that is potential of considerable relevance to L2 pronunciation teaching and learning practices.

Data availability statement. The datasets and R scripts for doing the analyses are available at OSF via the following link: <https://osf.io/he2am/>.

Acknowledgments. We acknowledge that part of the data from the Catalan speakers was collected by Mr. Xianqiang Fu at Universitat Pompeu Fabra. We sincerely thank the students at the Department of Translation and Language Sciences, Universitat Pompeu Fabra, and the students at Zhangqiu Experimental School (Jinan, China) who voluntarily participated in this study.

Competing interest. The authors declare no competing interests exist.

Funding. This study is funded by 'Multimodal Communication: The integration of prosody and gesture in human communication and in language learning' (PID2021-123823NB-I00) awarded by the Ministerio de

Ciencia e Innovación and ‘Multimodal language learning: Prosodic and Gestural Integration in Pragmatic and Phonological Development’ (PGC2018-097007-B-I00), awarded by the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación, and Fondo Europeo de Desarrollo Regional. P.L. is supported by the Research Council of Norway through its Centres of Excellence funding scheme (223265). F.B. acknowledges a Margarita Salas grant funded by the European Union-NextGenerationEU, Ministry of Universities and Recovery, Transformation and Resilience Plan, through a call from Pompeu Fabra University.

References

- Abdallah, F. (2010). *The role of phonological memory in L2 acquisition in adults at different proficiency levels*. [Doctoral dissertation, Université Laval]. <https://www.collectionscanada.ca/obj/thesescanada/vol2/QQLA/TC-QQLA-27300.pdf>
- Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2010). L2 speech learning in adulthood and phonological short-term memory. In K. Dziubalska-Kołaczyk, M. Wrembel, and M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 1–14). Poznań. <https://doi.org/10.2478/psicl-2011-0002>
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders* 36, 189–208. [https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)
- 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).
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- 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>
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLOS ONE* 8, e60676. <https://doi.org/10.1371/journal.pone.0060676>
- Boebinger, D., Norman-Haignere, S. V., McDermott, J. H., & Kanwisher, N. (2021). Music-selective neural populations arise without musical training. *Journal of Neurophysiology* 125, 2237–2263. <https://doi.org/10.1152/jn.00588.2020>
- Boll-Avetisyan, N., Bhatara, A., & Höhle, B. (2017). Effects of musicality on the perception of rhythmic structure in speech. *Laboratory Phonology* 8, 1–16. <https://doi.org/10.5334/labphon.91>
- Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2016). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Bilingualism* 19, 971–986. <https://doi.org/10.1017/S1366728915000425>
- Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning* 66, 774–808. <https://doi.org/10.1111/lang.12159>
- 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, 149–165. <https://doi.org/10.1177/0023830919826334>
- Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology* 32, 867–873. <https://doi.org/10.1037/0012-1649.32.5.867>
- Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences* 3, 923–940. <https://doi.org/10.3390/brainsci3020923>
- Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology* 4, 1–11. <https://doi.org/10.3389/fpsyg.2013.00874>
- 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, 169. <https://doi.org/10.3390/brainsci80901691>
- Christiner, M., Renner, J., Groß, C., Seither-Preisler, A., Benner, J., & Schneider, P. (2022). Singing Mandarin? What short-term memory capacity, basic auditory skills, and musical and singing abilities reveal about learning Mandarin. *Frontiers in Psychology* 13, 895063. <https://doi.org/10.3389/fpsyg.2022.895063>

- Christiner, M., Rüdegger, 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, 455–471. <https://doi.org/10.1080/14790718.2018.1424171>
- Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustical Society of America* 131, 4756–4769. <https://doi.org/10.1121/1.4714355>
- 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, 68. <https://doi.org/10.3390/languages8010068>
- 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, 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, 1–9. <https://doi.org/10.3389/fnins.2020.00618>
- Dolman, M., & Spring, R. (2014). To what extent does musical aptitude influence foreign language pronunciation skills? A multi-factorial analysis of Japanese learners of English. *World Journal of English Language* 4, 1–11. <https://doi.org/10.5430/wjel.v4n4p1>
- Duanmu, S. (2007). *The phonology of standard Mandarin*. Oxford: Oxford University Press.
- Eisenberg, I., Enkavi, Z., Bissett, P., Sochat, V., & Poldrack, R. (2017). Digit-span. Available at: <https://github.com/expfactory-experiments/digit-span>.
- Fortkamp, M. B. M. (2000). *Working memory capacity and L2 speech production: An exploratory study*. [Doctoral Dissertation, Universidade Federal de Santa Catarina]. <http://repositorio.ufsc.br/xmlui/handle/123456789/78287>
- François, C., Jaillet, F., Takerkar, S., and Schön, D. (2014). Faster sound stream segmentation in musicians than in nonmusicians. *PLoS ONE* 9, 0101340. <https://doi.org/10.1371/journal.pone.0101340>
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). irr: Various coefficients of interrater reliability and agreement version 0.84.1 [software]. Available at: <https://cran.r-project.org/package=irr>.
- Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *Journal of the Acoustical Society of America* 115, 2545.
- Jekiel, M., & Malarski, K. (2021). Musical hearing and musical experience in second language English vowel acquisition. *Journal of Speech, Language, and Hearing Research* 64, 1666–1682. https://doi.org/10.1044/2021_JSLHR-19-00253
- Jekiel, M., & Malarski, K. (2023). Musical hearing and the acquisition of foreign-language intonation. *Studies in Second Language Learning and Teaching* 13, 151–178. <https://doi.org/10.14746/ssl.t.23166>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism* 11, 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, 0052508. <https://doi.org/10.1371/journal.pone.0052508>
- Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience* 30, 7604–7612. <https://doi.org/10.1523/JNEUROSCI.0296-10.2010>
- Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America* 124, 3235–3248. <https://doi.org/10.1121/1.2990713>
- Li, M., & Dekeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition* 39, 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, 1015–1039. <https://doi.org/10.1017/S0272263120000054>
- Li, P., Zhang, Y., Fu, X., Baills, F., & Prieto, P. (2022). Melodic perception skills predict Catalan speakers' speech imitation abilities of unfamiliar languages. In S. Frota, M. Cruz, and M. Vigário (Eds.), *Proceedings of the 11th International Conference on Speech Prosody* (pp. 876–880). <https://doi.org/10.21437/SpeechProsody.2022-178>
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience* 23, 2701–2715. <https://doi.org/10.1162/jocn.2010.21585>
- Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience* 19, 1453–1463. <https://doi.org/10.1162/jocn.2007.19.9.1453>
- Martínez-Montes, E., Hernández-Pérez, H., Chobert, J., Morgado-Rodríguez, L., Suárez-Murias, C., Valdés-Sosa, P. A., et al. (2013). Musical expertise and foreign speech perception. *Frontiers in Systems Neuroscience* 7, 1–11. <https://doi.org/10.3389/fnsys.2013.00084>
- Milovanov, R., Huotilainen, M., Esquef, P. A. A., Alku, P., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience Letters* 460, 161–165. <https://doi.org/10.1016/j.neulet.2009.05.063>
- Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research* 1194, 81–89. <https://doi.org/10.1016/j.brainres.2007.11.042>
- Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: A review. *Frontiers in Psychology* 2, 1–6. <https://doi.org/10.3389/fpsyg.2011.00321>
- Mizera, G. J. (2006). Working memory and L2 oral fluency. Available at: <https://core.ac.uk/download/pdf/12207478.pdf>.
- Müllensiefen, D., Elvers, P., & Frieler, K. (2022). Musical development during adolescence: Perceptual skills, cognitive resources, and musical training. *Annals of the New York Academy of Sciences* 1518, 264–281. <https://doi.org/10.1111/nyas.14911>
- Murljadic, M. (2020). Musical ability and accent imitation. Available at: https://opencommons.uconn.edu/srhonors_theses/688.
- Navarro Pérez, P., & Rohrer, P. L. (2020). Digit-span. Available at: <https://github.com/pnavarro/digit-span>.
- Norman-Haignere, S., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., et al. (2022). A neural population selective for song in human auditory cortex. *Current Biology* 32, 1470–1484.e12. <https://doi.org/10.1016/j.cub.2022.01.069>
- Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. <https://doi.org/10.1016/j.neuron.2015.11.035>
- 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, 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, 557–581. <https://doi.org/10.1017/S027226310707043X>
- Ott, C. G. M., Langer, N., Oechslin, M. S., Meyer, M., & Jäncke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Frontiers in Psychology* 2, 1–10. <https://doi.org/10.3389/fpsyg.2011.00195>
- Pastuszek-Lipinska, B. (2008). Influence of music education on second language acquisition. *Journal of the Acoustical Society of America* 123, 3737–3737. <https://doi.org/10.1121/1.2935254>
- 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>.
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research* 308, 98–108. <https://doi.org/10.1016/j.heares.2013.08.011>
- Pei, Z., Wu, Y., Xiang, X., & Qian, H. (2016). The effects of musical aptitude and musical training on phonological production in foreign languages. *English Language Teaching* 9, 19. <https://doi.org/10.5539/elt.v9n6p19>

- Peretz, I., Vuvar, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions B* 370, 20140090. <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, 508–517. <https://doi.org/10.1177/0305735611415145>
- Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., *et al.* (2015). Intonational phonology of Catalan and its dialectal varieties. In S. Frota and P. Prieto (Eds.), *Intonation in romance* (pp. 9–62). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199685332.003.0002>
- R Core Team. (2014). R: A language and environment for statistical computing. Available at: <http://www.r-project.org/>.
- Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience* 31, 3843–3852. <https://doi.org/10.1523/JNEUROSCI.4515-10.2011>
- Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica* 138, 1–10. <https://doi.org/10.1016/j.actpsy.2011.03.007>
- Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences* 1252, 229–236. <https://doi.org/10.1111/j.1749-6632.2012.06447.x>
- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science* 17, 675–681. <https://doi.org/10.1111/j.1467-9280.2006.01765.x>
- 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, 259–280. <https://doi.org/10.1111/j.1467-9922.2010.00611.x>
- Voeten, C. C. (2021). *buildmer: Stepwise elimination and term reordering for mixedEffects regression*. Available at: <https://cran.r-project.org/package=buildmer>.
- Wheeler, M. W. (2005). *The phonology of Catalan*. New York: Oxford University Press.
- Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., *et al.* (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology* 33, 101–111. <https://doi.org/10.1080/13803395.2010.493149>
- 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, 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, 33–45. <https://doi.org/10.1111/nyas.13410>
- Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music* 48, 389–409. <https://doi.org/10.1177/0305735618804038>
- 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, 477–497. <https://doi.org/10.1177/0267658320978493>

Cite this article: Li, P., Zhang, Y., Baills, F., & Prieto, P. (2023). Musical perception skills predict speech imitation skills: differences between speakers of tone and intonation languages, *Language and Cognition*, 1–19. <https://doi.org/10.1017/langcog.2023.52>