



Reassessing the Determinants of English Pronunciation: Musical Aptitude vs. L1

Zhumageldina Aruzhan

aru.zhumageldina@gmail.com, Sophia University, Japan

APA Citation: Aruzhan, Z. (2026). Reassessing the determinants of English pronunciation: musical aptitude vs. L1. <i>LEARN Journal: Language Education and Acquisition Research Network</i> . 19(2), 473-500. https://doi.org/10.70730/IGHX2935	
Received 23/01/2026 Received in revised form 04/04/2026 Accepted DD/MM/YYYY (Garamond, 9pt)	ABSTRACT This study compared the effects of musical aptitude (MA) and native language (L1) on second language (L2) English segmental pronunciation proficiency. The uniqueness lies in the cross-linguistic analysis between native Russian, Mandarin Chinese, and Japanese EFL learners. Musical aptitude of the participants was assessed through Seashore Musical Aptitude Test, while the Pronunciation Test involved the Elicited Imitation Task with emphasis on segmental pronunciation performance, where the most problematic phonemes for the native speakers of each three language groups were included. Contrary to some prior research, results revealed a non significant to weak relationship between MA and pronunciation accuracy. Instead, L1 background was the dominant predictor, with a clear performance hierarchy (Russian > Chinese > Japanese) emerging regardless of MA scores. The findings suggest that the influence of MA is not universal but is augmented by musical training or the specific phonological relationship between the L1 and L2. A concurrent indication is that innate musical aptitude alone is insufficient for mastering L2 phonemes,

	underscoring the primacy of L1 transfer and the potential need for targeted pedagogical intervention over reliance on innate auditory talent. Keywords: musical aptitude, native language, English pronunciation, segmental pronunciation, Seashore Musical Aptitude Test.
--	---

Introduction

Accurate pronunciation is widely recognized as a central component of communicative competence in second language (L2) acquisition. A speaker's segmental accuracy has been repeatedly identified as a primary source of intelligibility breakdowns in L2 speech (Munro & Derwing, 2006). Poor segmental accuracy can impede mutual understanding in both everyday and professional contexts (Binmustafa & Aljarooshi, 2020), while greater accuracy facilitates clearer communication and supports speaker confidence (Islam, 2020).

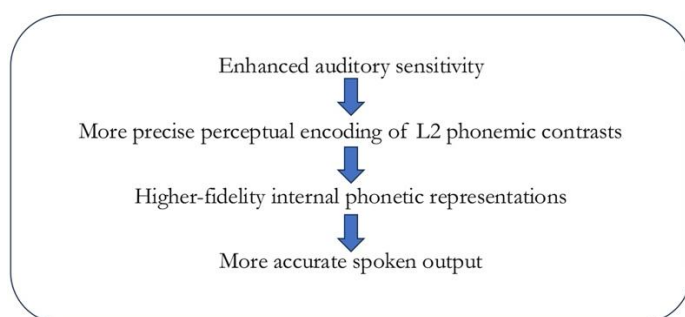
The theoretical basis for linking musical aptitude to L2 pronunciation lies in the relationship between auditory input processing and spoken output production. Speech production is not an independent skill; it is fundamentally shaped by the quality of a learner's perceptual representations. According to the Speech Learning Model (Flege, 1995) and the Perceptual Assimilation Model (Best, 1994), L2 learners can only accurately produce sounds they have first learned to perceptually distinguish from their L1 phonologically categorized equivalents. When two L2 sounds are assimilated to a single L1 category, the learner lacks a differentiated perceptual target, and production errors follow directly from this perceptual failure. The perception-production link is therefore not incidental but constitutive: what a learner hears and encodes internally determines the phonetic target they attempt to reproduce.

Musical aptitude enters this framework as a capacity that sharpens the precision of auditory input processing. The core components of musical aptitude such as pitch discrimination, tonal memory, and rhythmic sensitivity, rely on the same neural substrates involved in fine-grained phonological processing (Patel, 2011; Strait & Kraus, 2011). Learners with higher musical aptitude have been linked to greater sensitivity to subtle pitch and spectral contrasts relevant to speech perception (Kempe et al., 2015), as well as more accurate subcortical encoding of acoustic features and auditory working memory (Strait et al., 2011). These advantages are directly relevant to L2 phonological acquisition: a learner who can more precisely perceive acoustic distinctions between L2 sounds is better positioned to form differentiated

phonological representations, which in turn provide more accurate internal targets for production. The pathway from musical aptitude to pronunciation accuracy can therefore be understood from Figure 1:

Figure 1

Mediation Pathway from Musical Aptitude to Pronunciation Accuracy



Importantly, musical aptitude functions as an upstream perceptual resource rather than a direct motor skill. This theoretical pathway further predicts that musical aptitude should be most beneficial in tasks that rely heavily on perceptual encoding, such as controlled imitation, and potentially less so in spontaneous speech where automatized L1 motor routines may override perceptual input (Li et al., 2024; Zhumageldina, 2024).

The empirical literature on the music-L2 pronunciation relationship remains contested. Several studies have reported significant positive associations between musical aptitude and L2 segmental accuracy (Jekiel & Malarski, 2021; Milovanov et al., 2010; Slevc & Miyake, 2006), while others have failed to replicate these effects, attributing the variability to learner motivation, learning strategies, and L1-L2 phonemic overlap (Delogu et al., 2010; Zhumageldina, 2024). Direct comparison is further complicated by methodological differences – different aptitude tests, different pronunciation measures, and different learner populations – making it difficult to isolate the precise source of discrepancies.

This study directly addresses this methodological issue by examining the interplay of musical aptitude and L1 background on English segmental pronunciation of Japanese, Russian, and Chinese native speakers. The study focuses on these three languages for several reasons. In addition to representing distinct language families within the Expanding Circle, where English is primarily used for international communication (Islam, 2020), they represent typologically distinct phonological systems that each pose a unique set of challenges for English segmental acquisition. For instance, Japanese is

a mora-timed language with a small consonant inventory, severe restrictions on consonant clusters, and no phonemic /r/-/l/ distinction, leading learners to assimilate both sounds into a single native category (Iverson et al., 2003). Russian features pervasive palatalization contrasts, complex consonant clusters that do not map directly onto English permissible sequences, and a vowel reduction system that differs substantially from English. Mandarin, as a tonal language with predominantly CV syllable structure and highly restricted codas, creates a particular difficulty with English final consonants and clusters, while its segmental inventory lacks direct equivalents for several English contrasts, including dental fricatives and certain vowel distinctions (Swan & Smith, 2001). Together, these three L1 groups stress different dimensions of English segmental accuracy enabling a cross-linguistic comparison of how L1 phonology affects the interrelation between musical aptitude and pronunciation.

The study focuses on segmental accuracy for two reasons: first, to provide a direct comparison with prior conflicting studies which often used segmental measures; and second, because segmental errors are frequently identified as a primary source of intelligibility breakdowns (Munro & Derwing, 2006). The research is guided by two questions:

RQ1: What is the relationship between musical aptitude and L2 English segmental pronunciation accuracy across different L1 groups?

RQ2: How does the predictive strength of musical aptitude on L2 English segmental pronunciation accuracy compare to that of L1 background?

Literature Review

Musical Aptitude and its Assessment

Musical aptitude is often mistaken for musicianship, however, the two are distinct. The concept of “aptitude” itself is debated, as it does not necessarily develop through training. Schellenberg & Peretz (2008) define musical aptitude as a natural, untutored ability, whereas *musicianship* results from both aptitude and formal training. This distinction suggests that while musicians generally have high musical aptitude, individuals without training may still possess strong innate abilities. Seashore (1938) further argued that musical aptitude is inherited, with genetics determining its potential and environmental factors shaping musical achievement.

The measurement of musical aptitude is fraught with theoretical disagreement, primarily centering on whether it constitutes an innate, perceptual skill or a developed, cognitive one. This debate directly informs

the choice of musical assessment tool and its validity for language research. The most common standardized musical tests include:

The Seashore Musical Aptitude Test (SMAT), a standardized instrument that deconstructs music into isolated components – pitch, loudness, rhythm, timbre, time, and tonal memory – using artificially generated tones. Law & Zentner (2012) emphasize that SMAT comprehensively assesses key components of musical aptitude, making it a valuable tool for evaluating musical potential.

In contrast, Gordon's Primary and Advanced Measures of Music Audiation (PMMA / AMMA) and the Musical Aptitude Profile (MAP) represent a shift toward a more cognitive understanding of musical ability. They assess "audiation" – comprehension of music when the sound is no longer present – through tasks requiring participants to detect differences in melodic and rhythmic phrases. This aligns more closely with the processing of linguistic prosody. However, these tests are often age-specific and less frequently employed in second language acquisition (SLA) research, creating a methodological gap.

Wing's Standardized Tests sought a middle ground, incorporating harmonic and stylistic judgment thereby bridging perceptual and cognitive domains. However, its complexity has limited its widespread adoption in non-musical fields.

Relationship Between Musical Aptitude and L2 Pronunciation

Research has consistently identified a connection between auditory skills, musical aptitude, and L2 pronunciation proficiency, though this relationship is complex and context dependent. Empirical evidence supporting this link comes from multiple linguistic backgrounds. For instance, studies utilizing the SMAT have reported positive correlations for Finnish (Milovanov et al., 2008) and Spanish (Vangehuchten & Verhoeven, 2015; Rodríguez, 2021) learners, with findings indicating that those with heightened auditory skills excel in phonemic and stress pattern discrimination. This is further supported by research employing other metrics, such as the Musical Aptitude Profile, which has found supportive evidence with Polish (Jekiel & Malarski, 2021) learners.

The potential advantages of musical engagement are underscored by findings that Finnish choristers outperformed English philology students in both musical and English pronunciation tests (Milovanov et al., 2008, 2010), suggesting that musical training may bring specific benefits in linguistic contexts. The underlying cognitive mechanism for this synergy may involve enhanced auditory-motor integration and interhemispheric cooperation. Neuroscientific perspectives show that musical activities engage both cerebral

hemispheres – the left, specialized for linguistic processing, and the right, which handles musical and prosodic elements (Elliot, 1995). This simultaneous engagement optimizes learning by making the brain more receptive and active. For example, Saito et al. (2022) found that Japanese EFL learners who received auditory training showed greater improvement in both auditory and phonetic skills than those who received only phonetic training. Similarly, Shao et al. (2023) emphasized that the ability to remember and reproduce sounds is fundamental for achieving advanced pronunciation comprehensibility and nativelikeness.

However, this relationship is far from universal and is likely moderated by key methodological and learner-specific variables. The most compelling evidence for this complexity comes from a direct conflict within the literature that uses the same assessment tool. For example, while the SMAT revealed positive correlations in European contexts (Milovanov et al., 2008; Vanghechuten & Verhoeven, 2015), Dolman & Spring (2014) found no significant correlation between musical aptitude and segmental pronunciation among Japanese learners, despite using a similar pronunciation task to Milovanov et al. (2008). This contradiction cannot be dismissed as a mere artifact of different measurement instruments; it points to the powerful influence of other factors, such as L1 phonological background, that may overwhelm any advantage conferred by general auditory aptitude.

Relationship Between Mother Tongue and L2 Pronunciation

The influence of a learner's L1 on L2 pronunciation can be understood through cross-language perceptual similarity, which explains how learners perceive L1 and L2 sounds and how these perceptions impact their ability to accurately produce L2 sounds (Strange et al., 2011). For example, Japanese speakers may find it easier to pronounce the English /r/ sound over the /l/ sound because the English /r/ is more distinct from the Japanese /r/, whereas the English /l/ is closer to the Japanese liquid sound (Aoyama et al., 2004). This idea aligns with the Perceptual Assimilation Model (PAM), which suggests that challenges in L2 pronunciation arise from learners' difficulties distinguishing L1 and L2 sounds.

L1 influence can also cause interference in L2 pronunciation, particularly when L1 and L2 phonologies differ. This interference often leads to a “foreign accent”, especially when learners apply L1 patterns to L2 sounds, resulting in errors (Gass & Selinker, 2001). While some influence from L1 can be beneficial, it typically causes deviations from native-like pronunciation, making communication more challenging.

Orthographic differences also play a role in L2 pronunciation. Learners may apply their L1 spelling rules to L2 sounds, leading to

mispronunciations. For instance, English speakers learning Spanish might mispronounce the sounds /b/ and /v/ because of differences in spelling conventions (Zampini, 1994). Overall, the way learners process L2 sounds through the lens of their L1, shaped by perceptual similarity, interference, and orthographic influence, affects their ability to produce L2 sounds accurately.

Phonetic Challenges of Russian, Chinese, and Japanese EFL Learners

The segmental pronunciation errors common to Russian, Chinese, and Japanese learners of English are not random; they are systemic with predictable consequences due to L1 phonological filtering. PAM and SLM provide a framework for understanding why these specific phonemic categories pose such persistent challenges.

a. Native Russian (NR) Learners: Articulatory Setting and Phonemic Gaps

NR speakers face challenges rooted in both articulatory setting and phonemic gaps. A primary challenge is the English /ɹ/, which differs from the Russian alveolar trill /r/, leading to substitutions that affect intelligibility (Bergman, 2021). Unlike the retroflex motion of the English /ɹ/, Russian speakers produce a more retracted trill, making it difficult to approximate the English sound (Islam, 2020). Similarly, they frequently substitute the dental fricatives /θ/ and /ð/ with /s/, /d/, or /z/ due to their absence in Russian, resulting in mispronunciations such as ‘thin’ as ‘sin’ or ‘then’ as ‘zen’ (Bergman, 2021). Another common issue is the /v/–/w/ distinction, where Russian speakers may interchange these sounds, leading to errors like ‘west’ pronounced as ‘vest’ or conversely (Swam & Smith, 2001).

b. Native Mandarin Chinese (NMC) Learners: Prosodic Interference and Category Formation

For NMC speakers, the primary challenges often stem from prosodic or vocalic L1-L2 mismatch. Mandarin is a tonal language where pitch variation occurs at the syllable level to signal lexical meaning, not at the phrase level for intonation as English. This can lead to the misapplication of tonal patterns onto English segments. Moreover, they face difficulties with vowel length contrasts, particularly distinguishing between [i] and [i:], as Mandarin lacks a phonemic distinction based on vowel duration (Han, 2013). This often results in errors like ‘ship’ and ‘sheep’ being pronounced identically. Another persistent issue is the replacement of English dental fricatives /θ/ and /ð/ with /t/, /d/, or /s/ due to the absence of these sounds in their L1 (Fang, 2022). Finally, similar to NR learners, NMC tend to replace /v/ with /w/,

leading to pronunciations like ‘vet’ as ‘wet’, a challenge that persists even among advanced learners (Zhang et al., 2021).

c. *Native Japanese (NJ) Learners: The Single-Category Assimilation Problem*

The most documented challenge for NJ speakers is the /r/ and /l/ contrast, a classic example of PAM-L2’s Single-Category Assimilation. Both English phonemes are perceptually mapped onto the single Japanese liquid /r/, making them exceptionally difficult to discriminate (e.g., “royal” and “loyal” sound the same when pronounced by many NJ learners) (Wayland, 2021). They also tend to compress long vowels, such as /i:/, due to differences in syllabic structure between English and Japanese (Lee, 2018). Lastly, similar to NR and NMC groups, NJ speakers replace dental fricatives /θ/ and /ð/ with /s/ and /z/, respectively, leading to mispronunciations like ‘think’ as ‘sink’ (Saito, 2011).

Table 1

Problematic English Phonemes for Russian, Mandarin Chinese, and Japanese Native Speakers Chosen for This Study

Russian	Mandarin Chinese	Japanese
	fricatives [θ] and [ð]	
	/w/ and /v/	/r/ and /l/
/ɹ/ replaced with trill /r/	long vowel [ɪ:] and short vowel [ɪ]	

The three areas discussed above (musical aptitude, native language, L2 pronunciation) together provide the theoretical foundation for the present study. Understanding musical aptitude and how it is measured is essential before examining its role in L2 pronunciation, as the choice of assessment tool directly shapes what conclusions can be drawn. While research has shown that learners with stronger musical aptitude tend to perform better in L2 pronunciation tasks, this relationship is not consistent across all learner groups. This inconsistency points to the important role of L1 background, as learners perceive and produce L2 sounds through the filter of their native phonological system. The specific challenges faced by NR, NMC, and NJ learners further illustrate how L1 structure shapes L2 pronunciation errors in predictable ways. Understanding how musical aptitude and L1 phonological background interact across different language groups is essential to the aims of the present study.

Methodology

Participants

A total of 46 participants were recruited for this study, comprising native speakers of Russian, Mandarin Chinese, and Japanese languages (17, 14, and 15 respectively). The sample size reflects the maximum number of eligible volunteers available within the recruitment constraints of the study; while larger samples are generally preferable for statistical power, this number is consistent with those employed in comparable experimental phonetics and music-language aptitude research (e.g., Milovanov et al., 2010), and is considered sufficient for the exploratory purposes of the present investigation.

All participants were enrolled in bachelor’s or master’s degree programs at universities in Kazakhstan and Japan. All of them were in the same age bracket of 18-25 years and had a self-reported CEFR English proficiency level of at least A2. The key control criteria were that no participant had a prior experience of English pronunciation training, lived in an English-speaking country nor had interactions with a native English speaker for longer than 2 weeks, ensuring that their pronunciation was primarily influenced by their L1 and not by prolonged immersion. Another condition was that participants had not received any musical training to ensure the performance on the SMAT reflected innate musical abilities.

Table 2

Participants’ Profile

		F	%	N
Gender	Male	20	45%	46
	Female	26	55%	
Native Language	Japanese	17	38%	46
	Mandarin Chinese	14	30%	
	Russian	15	32%	
Academic Stage	Bachelor’s degree	34	74%	46
	Master’s degree	12	26%	

Musical Aptitude Test

The Seashore Musical Aptitude Test, a standardized test battery for assessing musical abilities, was utilized to evaluate participants’ musical aptitude. While other instruments, such as Gordon’s Advanced Measures of Musical Audiation (AMMA) are valuable for assessing higher-level cognitive aspects of musicality, often termed “audiation” or the internal

comprehension of musical structure, the SMAT was chosen for its unique ability to isolate and measure low-level auditory acuity (the ability to perceive subtle sound differences).

The core premise of the Shared Mechanisms Hypothesis, which this study seeks to test, posits that domain-general auditory processing abilities underlie both musical and speech perception. The SMAT directly operationalizes this construct by deconstructing music into its fundamental psychoacoustic components: pitch, time (duration), timbre, loudness, rhythm, and tonal memory. For the pitch, time, timbre, loudness, and rhythm subtests, participants were tasked with identifying whether paired notes were the same or different. In the tonal memory subtest, they were required to detect discrepancies within sequences of three to five tones (see Appendix). By utilizing artificially generated tones in a “same/different” discrimination task, the test minimizes the influence of prior cultural exposure or cognitive interpretation of melody and harmony.

Participants accessed the test through a link embedded in a private Google Form. They were instructed to complete the test in a single session in a quiet setting using headphones to ensure sound clarity and minimize distractions. Their responses were automatically recorded within the Google Form, ensuring privacy as only the primary researcher had access to the results.

Pronunciation Test

A customized Elicited Imitation Task (EIT) was developed to assess segmental pronunciation proficiency in the Pronunciation Test (PT) stage. The EIT addressed the most challenging English phonemes for Russian, Chinese, and Japanese EFL learners. The test consisted of 30 sentences. The target words within these sentences were carefully selected to include English phonemes known to be problematic for native speakers of the three languages, based on established Contrastive Analysis frameworks (Table 3).

Each sentence was designed to be syntactically simple and semantically unambiguous to minimize cognitive load, ensuring that pronunciation performance was the primary measured variable. Although the target phonemes were specific to each L1 group, all participants pronounced all 30 sentences. This provided a consistent and comprehensive dataset, allowing for a direct comparison of how different L1 backgrounds handle the same set of phonetic challenges.

Table 3

Sentences Used in the Pronunciation Test

Russian	His face was wet with tears.
	I'm going to take a bath .
	Let's meet at the Central Park .
	I th ought it was a joke.
	I haven't talked to my moth er in years.
	This is not the ri ght time.
	Peel it carefully.
	Take the pi ll.
	We must save the w orld.
Chinese	He was walking down the path .
	We live in the village .
	May I open the w indow?
	Sheep are grazing here.
	I'll pack and sh ip it.
	The weath er continues hot.
	We drove th rough the tunnel.
	I'll w ait until tomorrow.
	He is the vice -president.
Japanese	The phone sl ipped out of my hands.
	I need some sleep .
	Just dip it in the sauce.
	Just take a deep breath.
	Nei ther of them is good.
	She made grass grow.
	She broke the glass .
	I miss my paren ts.
	She had a thin face.
	The silence was thick .
	He's so tall and lean.
	You can choose oth er shoes.

Data Processing

The submitted audio recordings were first anonymized by removing any identifying metadata and assigning each participant a random numerical code. To ensure unbiased evaluation, the files were then randomized and presented to three trained linguists for blind rating. The raters, all native English speakers with phonetic training, were unaware of the participants' native language and musical aptitude scores. They evaluated the segmental accuracy of the 30 target phonemes for each participant using a standardized scale (e.g., 1 to 3, where 1 = production bears little resemblance to the target, 2 = production approximates the target but contains noticeable inaccuracies,

3 = production closely matches the native speaker model). Each phoneme was thus scored independently against a native speaker model recording.

The decision to employ a native-likeness scale rather than an intelligibility-based measure was deliberate and theoretically motivated. The present study was not concerned with whether participants could communicate successfully in English, but rather with the degree to which they could accurately perceive and reproduce specific target phonemes modelled by a native speaker. This focus on phonetic imitation accuracy is directly relevant to the study's research questions concerning L1 phonological transfer and musical aptitude, both of which operate at the level of sound production rather than communicative function.

Reliability and Validity of the Pronunciation Test

Cronbach's alpha was calculated to assess reliability across three language groups, with all of them showing moderate inter-rater reliability coefficient (Table 4).

Table 4

Internal Consistency Reliability of the Pronunciation Test by L1 Group

Group	Cronbach's alpha
Russian	.759
Chinese	.799
Japanese	.700

A content validity test was conducted with five expert raters (three Russian, one Chinese, and one Japanese English teachers) to ensure the test sentences represented the phonetic challenges relevant to each language group. The Content Validity Index (CVI) was calculated at both the item and scale level (Table 5). The average item-level CVI (S-CVI/Ave) was .82, suggesting strong agreement among raters regarding the relevance of the test items. Additionally, 53% of the items received unanimous agreement (S-CVI/UA = .53), reflecting a moderate level of consensus on their relevance. These results support the content validity of the pronunciation test, confirming it measures the intended phonetic features effectively.

Table 5

Content Validity Index of the Pronunciation Test

CVI Type	Value
S-CVI/Ave	.82
S-CVI/UA	.53

Results

Assumption Testing

Prior to the main analysis, the assumptions underlying the two-way ANOVA and linear regression were examined. Normality of distribution within each L1 group was assessed using the Shapiro-Wilk test (Table 6). Results confirmed that both pronunciation scores and musical aptitude scores were normally distributed across all groups. Although the Chinese group's pronunciation scores approached the conventional threshold ($p = .076$), they remained within the acceptable range. Homogeneity of variance was assessed using Levene's test (Table 7), which indicated equal variances across groups for both pronunciation scores ($F(2, 43) = 0.523$, $p = .597$) and musical aptitude scores ($F(2, 43) = 1.899$, $p = .162$). All assumptions were therefore satisfied, confirming the appropriateness of the parametric analysis reported below.

Table 6

Normality Test of Pronunciation (PT) and Musical Aptitude (MA) Scores

		PT		MA	
L1 Group	df	Statistic	Sig.	Statistic	Sig.
Russian	17	.959	.679	.984	.989
Chinese	14	.888	.076	.912	.167
Japanese	15	.940	.317	.934	.250

Table 7

Homogeneity of Variances Test of Pronunciation and Musical Aptitude Scores

	Levene Statistic	Sig.
Pronunciation (PT)	0.523	.597
Musical Aptitude (MA)	1.899	.162

Descriptive Statistics

The distribution of average scores in SMAT and the Pronunciation Test (Table 8) demonstrates that Russian participants achieved the highest average PT score (75.2). Paradoxically, they also recorded the lowest overall musical aptitude score (192.6). This immediately suggests that excelling in English pronunciation does not necessarily depend on high scores in this particular test of musical ability. The Chinese group presented a different profile. They scored highest overall in the SMAT (212.5) and came in second in the Pronunciation Test (65.7), showing strong musical and auditory skills. The Japanese group, however, adds a further layer of complexity. They demonstrated solid musical aptitude (209.47), performing nearly as well as the Chinese group, yet they obtained the lowest pronunciation scores by a significant margin (56.4). This disconnect indicates that having a good “musical ear,” as measured by the SMAT, did not translate into more accurate English segmental pronunciation for these learners.

Table 8

Distribution of Average Scores in SMAT and Pronunciation Test (N = 46)

	<i>k</i>	Russian	Mandarin	Japanese
Loudness	50	34.2	39.35	39.82
Time	50	38.53	41.85	45.17
Timbre	50	35.86	37.57	31.88
Pitch	50	37.26	43.64	42.41
Rhythm	30	25.8	26.14	26.17
Tonal Memory	30	20.93	23.92	24
Overall SMAT score	260	192.6 (SD: 33.34)	212.5 (SD: 23.86)	209.47 (SD: 17.54)
Pronunciation Test	90	75.2 (SD: 7.66)	65.7 (SD: 9.66)	56.4 (SD: 7.25)

Note. k indicates the maximum possible score in each test.

The Relationship Between Musical Aptitude and Pronunciation Accuracy

To formally test the link between musical aptitude and pronunciation, a linear regression was performed with MA as a predictor, and PT as the outcome (Table 9). It found no significant relationship between MA and PT performance ($p = .654$). The model estimated an intercept of 71.32 ($p < .001$), indicating that in the absence of any influence from Musical Aptitude, the predicted pronunciation score would be approximately 71.32.

Table 9

Results of the Single Linear Regression Analysis Examining the Effects of MA on PT

Variable	B	SEB (std error)	β (std beta)	t-value	p-value
Intercept	71.32	13.24		5.39	< .001
Musical Aptitude	-0.028	0.06	-.00383	-0.45	.654

Note. $R^2 = .0046$

The Relationship Between Native Language and Pronunciation Accuracy

Given the weak role of musical aptitude, we next investigated the influence of the participants' native language. A multiple linear regression analysis (Table 10) was carried out to analyze the effects of three language groups on PT performance. The model was highly significant, with L1 background alone explaining a substantial 49.5% of the variance in pronunciation scores ($R^2 = .4948$). This shows that L1 background was a powerful and highly significant predictor of pronunciation performance.

Using the Russian speakers as the baseline, the analysis quantified the L1 effect: Chinese speakers scored, on average, 9.41 points lower than Russian speakers ($p = .003$). The effect was even more pronounced for Japanese speakers, who scored 18.79 points lower ($p < .0001$).

Table 10

Results of the Multiple Linear Regression Analysis Examining the Effects of L1 on PT

Predictor	B	SEB (std error)	β (std beta)	t-value	p-value
Intercept	75.200	2.111		35.62	< .001
L1 (Chinese vs. Russian)	-9.414	3.039	-1.73	-3.098	.003
L1 (Japanese vs. Russian)	-18.788	2.897	-3.45	-6.486	< .001

Note. $R^2 = .4948$

Interaction Between Native Language and Musical Aptitude

To bring the threads together, a Two-Way ANOVA (Table 11) was conducted, which confirmed a strong main effect of L1 on pronunciation scores ($F(2,40) = 20.88$, $p < .001$). In contrast, no significant effect of MA ($F(1,40) = 1.29$, $p = .263$) or interaction between L1 and MA ($F(2,40) = 0.67$,

$p = .517$) was indicated. This final test solidifies the narrative: a learner's native language is a dominant and independent predictor of their segmental pronunciation accuracy in English. Moreover, this suggests that while MA may contribute to certain aspects of language learning, it does not directly influence segmental pronunciation accuracy, nor does it differentially affect pronunciation performance across Russian, Chinese, or Japanese language groups.

Table 11

Results of the Two-Way ANOVA Analysis on the Differences in the Effects of L1 and MA on PT Across Groups

Source	df	Sum Sq	Mean Sq	F value	P-value	η_p^2
L1	2	2816.1	1408.0	20.881	< .001	.511
MA	1	87.1	87.1	1.291	.263	.031
L1:MA	2	90.6	45.3	0.672	.517	.033
Residuals	40	2697.2	67.4			

Phonetic Analysis of L1 Transfer Effects

Finally, Table 12 reveals significant differences in the pronunciation of problematic phonemes across Russian, Chinese, and Japanese speakers, likely influenced by the phonetic structures of their native languages. Table 12 represents the average scores out of perfect mean score of 3.00 for the pronunciation of each phoneme by each of the language groups. The results highlight the common struggle learners face with dental fricatives, which are absent in the phonetic systems of these languages.

Table 12

Mean Pronunciation Scores of Problematic Phonemes by L1

Phoneme	Russian Group	Mandarin Group	Japanese Group
Short /i/ Vowel	3.00	2.79	2.41
Long /i:/ Vowel	2.86	1.57	1.76
/r/	2.71	2.76	2.47
/l/	2.57	2.24	1.47
/w/	2.36	2.29	1.76
/v/	2.43	2.14	1.50
[ð]	2.43	1.71	1.35
[θ]	1.86	1.71	1.35

Note. The maximum score for each phoneme is 3.00.

Discussion

The present study aimed to investigate the relationship between musical aptitude (MA), first language (L1), and English segmental pronunciation performance among Russian, Chinese, and Japanese native speakers. Contrary to expectations based on prior research that found positive relationship between MA and English pronunciation (e.g., Christiner & Reiterer, 2013; Coumel et al., 2023; Milovanov et al., 2008, 2010; Saito et al., 2022), the relationship in this study was weak or even negative in some cases, while L1 demonstrated a stronger influence.

The most striking finding is the clear performance hierarchy in the PT with the Russian participants outperforming their Chinese and Japanese counterparts. This finding cannot be explained by musical aptitude, as the Russian group had the lowest overall SMAT score. This dissociation challenges the assumed centrality of general auditory acuity in L2 segmental acquisition and redirects focus toward L1 phonological transfer. In terms of sociolinguistic context, Russian and Japanese groups were recruited within their native or effectively native environments. The Mandarin-speaking group represents the one exception, as they were the only group participating from a country where their L1 has no societal presence. Notably, if immersion in a non-native environment exerts any facilitative effect on language processing, this would if anything be expected to benefit the Mandarin group, yet they did not outperform the Russian group. This confirms that the environmental exposure does not predict language control mechanisms (Coumel et al., 2026), attributing the performance hierarchy to L1 phonological factors.

Nevertheless, several alternative explanations for the observed performance hierarchy warrant consideration. First, although all participants self-reported an A2 level of English proficiency, this criterion is relatively broad and self-assessed, meaning that subtle variations in actual proficiency within and across groups cannot be entirely ruled out as a contributing factor. Thus, the relationship between proficiency level and musical aptitude is an interesting avenue for future research, particularly in designs that incorporate standardized L2 proficiency measures. Second, cross-national differences in EFL instructional quality, methodology, and cumulative contact hours with English may have differentially shaped pronunciation development across the three groups in ways that were not systematically controlled. Finally, motivational and attitudinal differences toward English learning may have contributed to differential levels of engagement and effort during the task. While L1 phonological transfer remains the most theoretically motivated explanation, these alternative factors should be considered in future research through more explicit controls for proficiency, instructional background, and learner-related variables.

No significant relationship was found between SMAT scores and pronunciation accuracy. This indicates that auditory acuity, in isolation, is an insufficient predictor of L2 segmental success. One plausible explanation concerns the nature of what the SMAT specifically measures: the ability to discriminate tones produced by musical instruments according to predetermined notes. These instrumental tones do not share the same acoustic and articulatory properties as human speech sounds, which are produced through vocal imitation and shaped by complex motor-articulatory processes. Pronunciation ability is likely tied to one's capacity to perceive and reproduce speech-specific sounds, a skill that instrument-tone discrimination does not directly tap into. This is not to suggest that musicality and speech processing are entirely unrelated because prior studies have, indeed, reported significant associations. Instead, these results raise questions about the construct validity of the SMAT as a proxy for musicality in L2 pronunciation research. Its reliance on instrumental tone discrimination, absent any vocal or articulatory component, may systematically underestimate the music-speech link by measuring only passive tonal acuity rather than the active, speech-relevant dimensions of musicality. Studies that found positive relationships between musicality and pronunciation largely drew on participants who received prior formal music training (e.g., Milovanov et al., 2010) or SMAT-based musical training over the course of the experiment (e.g., Borodkin, 2022). These experiences likely cultivated transferable skills such as heightened meta-awareness of sound and improved auditory working memory that go well beyond the passive tonal discrimination assessed by the SMAT.

Beyond instrument choice, the weak MA-pronunciation relationship may also reflect the domain-specific nature of the cognitive processes involved in L2 segmental acquisition. Domain-general auditory processing abilities, of the kind tapped by the SMAT, may be insufficient on their own to support the highly specialized task of acquiring L2 phonemic categories. Despite sharing the same neurological and cognitive processes (Brown et al., 2006; Kunert et al., 2015), L2 segmental acquisition requires not only the detection of acoustic differences but also the restructuring of existing phonological representations, a process that draws on language-specific rather than general auditory mechanisms. This perspective is consistent with models of speech perception that emphasize the role of phonological working memory, articulatory planning, and category formation as distinct from general auditory sensitivity. It is therefore possible that domain-general musical aptitude, while potentially facilitative under certain conditions, does not engage the language-specific processing pathways sufficiently to produce measurable gains in segmental production accuracy.

Moreover, the paradox of the Japanese group – high in low-level acuity but low in pronunciation scores – exemplifies that musical aptitude may only be activated as a significant predictor when the L1-L2 phonological distance is moderately challenging but not prohibitive. For this group, the task of differentiating English /r/ and /l/ requires more than fine acoustic discrimination; it demands establishment of new phonemic categories, a high-level cognitive task that appears to be beyond the purview of the auditory skills measures by the SMAT. This finding suggests that while basic auditory processing may be shared, its utility is powerfully constrained by the L1 phonological filter, which aligns with the foundational L2 speech learning theories. Flege et al.'s (1995) Speech Learning Model and Best's (1994) Perceptual Assimilation Model both posit that phonetic similarity or divergence between L1 and L2 phonemes affect how learners perceive and produce L2 sounds which was also corroborated by recent studies by Nagle et al. (2023) and Shafer et al. (2021).

The perception-production relationship, in its turn, is another boundary condition of this study. The present study operationalized pronunciation ability through spoken production, yet the auditory mechanisms assessed by the SMAT map more directly onto perceptual tasks. A dissociation between MA-perception and MA-production relationships is theoretically plausible and remains empirically unresolved. Future research incorporating both modalities within the same design would clarify whether musical aptitude facilitates the perceptual encoding of L2 phonemes even when it does not translate into measurable production gains.

Finally, the pronunciation task was restricted to a small set of phonemes selected to enable direct cross-group comparison based on known L1-specific difficulties. While this decision was methodologically motivated, the results may not extend to other English segmental features not included in the task. Their segmental position – that is, whether target phonemes appeared in word-initial, medial, or final position – requires control. Since segmental position is known to influence both the acoustic realization of phonemes and the relative difficulty of their production, future research should systematically account for positional effects in task design.

Overall, the discrepancy between the present findings and prior studies reporting significant MA-pronunciation relationships must be interpreted within methodological reflection. First, many prior studies included participants with formal musical training, which confounds innate aptitude with cultivated perceptual and motor skills. This study, by contrast, recruited only participants with no musical training at all, allowing for the isolation of raw, untutored auditory potential as measured by the SMAT, suggesting that the positive MA-pronunciation relationships reported elsewhere may reflect the effects of musical training rather than aptitude, which aligns with

Borodkin (2022). Second, prior studies examining learners whose L1 is phonologically closer to English – such as Finnish (Milovanov et al., 2008) or Spanish (Rodríguez, 2021; Vangehuchten & Verhoeven, 2015) – may find stronger MA effects precisely because the phonological filter is less obstructive. However, when L1-L2 phonological distance is very large, the role of general auditory aptitude may be overshadowed by the demands of phonemic category restructuring, which requires language-specific rather than domain-general processing, as suggested by both SLM and PAM (Flege et al., 1995; Best, 1994). Third, differences in outcome measures across studies further complicate direct comparison. Suprasegmental features such as stress, syllabic durations or intonation, are more directly linked to musical sensitivity than the segmental accuracy measured in the present study (Saito et al., 2019).

Finally, the extent to which confounding variables such as immersion experience, proficiency level, and the amount of L2 exposure were controlled varies considerably across the literature, and their uncontrolled presence in prior studies may have produced spurious or inflated correlations. Taken together, these methodological differences offer a coherent and evidence-grounded explanation for why the present findings diverge from parts of the existing literature.

Conclusion

In summary, this study contributes to the growing body of research on the relationship between musical aptitude, native language, and L2 pronunciation. While the findings highlight the significant role of L1 phonology in shaping pronunciation performance, they challenge the assumption that MA directly enhances segmental pronunciation accuracy. The weak or absent MA-pronunciation relationship observed here is interpreted not as evidence against a music-language connection broadly, but as a reflection of the limitations of untrained auditory aptitude as a predictor, the domain-specific cognitive domains of L2 phoneme acquisition, and the constraining role of L1 phonological filter. Taken together, these findings refine the scope of the Shared Mechanisms Hypothesis and underscore the importance of distinguishing between innate musical aptitude and the transferable skills cultivated through formal musical training.

In addition to addressing the limitations, future longitudinal studies could investigate whether musical training or aptitude development over time has a delayed or cumulative effect on pronunciation accuracy. Moreover, incorporating advanced analytical tools, such as structural equation modeling (SEM) or neuroimaging techniques (e.g., fMRI or EEG), could provide deeper insights into the cognitive and neural mechanisms underlying the MA-pronunciation relationship.

From a pedagogical perspective, the findings highlight the importance of tailoring pronunciation instruction to address L1-specific challenges. For example, Japanese learners may benefit from targeted training on /r/, /w/, and /v/ sounds, while all learners could benefit from focused practice on dental fricatives. Additionally, technology-enhanced tools, such as Automatic Speech Recognition software (e.g. ELSA Speak) could provide real-time feedback and immersive practice opportunities for learners. Such software, especially complemented with peer- and/or teacher-correction, is likely to yield positive outcomes.

About the Author

Zhumageldina Aruzhan: A doctoral student at the Graduate School of Languages and Linguistics, Sophia University, Japan.

References

- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, 32(2), 233-250. [https://doi.org/10.1016/S0095-4470\(03\)00036-6](https://doi.org/10.1016/S0095-4470(03)00036-6)
- Bergman, D. (2021). *Segmental pronunciation and error gravity in L2 Russian speakers of English: A data-driven approach* (Master's thesis, Itä-Suomen yliopisto). <https://erepo.uef.fi/handle/123456789/26672>
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, 167(224), 233-277. <https://www.researchgate.net/publication/243772232>
- Binmustafa, A.A. & Aljarooshi, N.A. (2020). Exploring pronunciation errors of primary pupils' school in Misrata Libya. *Scientific Journal of Faculty of Education, Misurata University-Libya*, 1(15), 193-205. <https://journals.misuratau.edu.ly/edu/upload/file/R-953-09%20Exploring%20Pronunciation%20Errors%20of%20primary%20pupils%E2%80%99%20school%20in%20Misrata%20Libya..pdf>
- Borodkin, K., Gassner, T., Ershaid, H., & Amir, N. (2022). tDCS modulates speech perception and production in second language learners. *Scientific Reports*, 12(1), 16212. <https://doi.org/10.1038/s41598-022-20512-0>

- Brown, S., Martinez, M., & Parsons, L. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, 23(10), 2791–2803. <https://doi.org/10.1111/j.1460-9568.2006.04785>
- 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>
- 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>
- Coumel, M., Liu, C., Trenkic, D., & de Bruin, A. (2026). Language control adapts to the immediate but not to the overall language environment during language switching in production and comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0001591>
- 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>
- 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(4), 1. <https://doi.org/10.5430/wjel.v4n4p1>
- Elliott, A. R. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal*, 79(3), 356-371. <https://doi.org/10.1111/j.1540-4781.1995.tb01112.x>
- Fang, A. L. S. (2022). Common pronunciation errors of Chinese-speaking students in Malaysia. *Higher Education and Oriental Studies*, 2(1). <https://doi.org/10.54435/heos.v2i1.42>
- Flege, J. E. (1995). Second language speech learning theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research/York*. <https://www.researchgate.net/profile/James-Flege/publication/333815781>
- Gass, S. M., & Selinker, L. (2001). *Second language acquisition. An introductory course* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates. https://www.researchgate.net/profile/Larry-Selinker/publication/229078359_Second_Language_Acquisition_An_Introductory_Course/links/5592ae3c08ae1e9cb4296f95/Second-Language-Acquisition-An-Introductory-Course.pdf

- Han, F. (2013). Pronunciation problems of Chinese learners of English. *Ortesol Journal*, 30, 26-30.
<https://eric.ed.gov/?id=EJ1152473>
- Islam, S. M. (2020). Segmental errors in English pronunciation of non-native English speakers. *Journal of Education and Social Sciences*, 16(1), 14-24. <https://www.jesoc.com/wp-content/uploads/2020/12/JESOC16-011.pdf>
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-B57.
[https://doi.org/10.1016/s0010-0277\(02\)00198-1](https://doi.org/10.1016/s0010-0277(02)00198-1)
- 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_JSLHR-19-00253
- Kempe, V., Bublitz, D., & Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology*, 106(2), 349-366. <https://doi.org/10.1111/bjop.12092>
- Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLOS ONE*, 10(11).
<https://doi.org/10.1371/journal.pone.0141069>
- Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PloS one*, 7(12), e52508.
<https://doi.org/10.1371/journal.pone.0052508>
- Lee, J. K. Y. (2018). Volleyball or Bareebooru? Common problems of English pronunciation for Japanese learners. *Gifu daigaku Kyoikusuishin. Gakusei shien kiko nenpo*, 4. https://www1.gifu-u.ac.jp/~orphess/nenpou/nenpou/Documents/2018nenpo_104.pdf
- Li, P., Zhang, Y., Bails, 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/langcog.2023.52>
- Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. 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., Pietilä, P., Tervaniemi, M., & Esquef, P. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, 20(1), 56-60. <https://doi.org/10.1016/j.lindif.2009.11.003>
- Munro, M. J., & Derwing, T. M. (2006) The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520-531. <https://doi.org/10.1016/j.system.2006.09.004>
- Nagle, C., Bruun, S., & Baese-Berk, M. M. (2023, July). Perceptual assimilation of Spanish, Thai, and Korean stop consonants by native English speakers. In *International Congress of Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/890.pdf
- Patel, A. D. (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>
- Rodríguez, A. P. (2021). Musical aptitude and foreign language receptive pronunciation. *Phonica*, 17, 72-89. <https://doi.org/10.1344/phonica.2021.17.72-89>
- Saito, K. (2011). Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English. *RELJ Journal*, 42(3), 363-378. <https://doi.org/10.1177/0033688211420275>
- Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, 22(5), 1123-1140. <https://doi.org/10.1017/S1366728918000895>
- Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Incidental and multimodal high variability phonetic training: Potential, limits, and future directions. *Language Learning*, 72(4), 1049-1091. <https://doi.org/10.1111/lang.12503>
- Schellenberg, E. G., & Peretz, I. (2008). Music, language and cognition: unresolved issues. *Trends in cognitive sciences*, 12(2), 45-46. <https://doi.org/10.1016/j.tics.2007.11.005>
- Seashore, C. E. (1938). The psychology of music. *Music Educators Journal*, 23(4), 30-33. <https://www.ibiblio.org/guruguha/MusicResearchLibrary/Books-English/BkE-Seashore-Psychologyofmusi030417mbp-0076.pdf>
- Shafer, V. L., Kresh, S., Ito, K., Hisagi, M., Vidal, N., Higby, E., ... & Strange, W. (2021). The neural timecourse of American English vowel discrimination by Japanese, Russian and Spanish second-

- language learners of English. *Bilingualism: Language and Cognition*, 24(4), 642-655.
<https://doi.org/10.1017/S1366728921000201>
- Shao, Y., Saito, K., & Tierney, A. (2023). How does having a good ear promote instructed second language pronunciation development? Roles of domain-general auditory processing in choral repetition training. *TESOL Quarterly*, 57(1), 33-63.
<https://doi.org/10.1002/tesq.3120>
- 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.1467-9280.2006.0176>
- Strait, D., & Kraus, N. (2011). Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. *Music perception*, 29(2), 133-146. <https://doi.org/10.1525/mp.2011.29.2.133>
- Strange, W., Hisagi, M., Akahane-Yamada, R., & Kubo, R. (2011). Cross-language perceptual similarity predicts categorical discrimination of American vowels by naïve Japanese listeners. *The Journal of the Acoustical Society of America*, 130(4), EL226-EL231
<https://doi.org/10.1121/1.3630221>
- Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems* (Vol. 1). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511667121>
- Vangenhuchten, L., Verhoeven, V., & Thys, P. (2015). Pronunciation proficiency and musical aptitude in Spanish as a foreign language: results of an experimental research project. *Revista de Lingüística y Lenguas Aplicadas*, 10, 90-100.
<https://doi.org/10.4995/rlyla.2015.3372>
- Wayland, R. (Ed.). (2021). *Second language speech learning*. Cambridge University Press. <https://doi.org/10.1017/97811088886901>
- Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania*, 470-481. <https://www.jstor.org/stable/344974>
- Zhang, B., Zhang, J., & Lee, S. (2021). Perception and production of English fricatives by Chinese learners of English: Error patterns and perception-production relationship. *The Korean Society of Speech Sciences*, 13, 25-36. <https://doi.org/10.13064/ksss.2021.13.1.025>
- Zhumageldina, A. (2024). Effect of musical aptitude on suprasegmental pronunciation in English: A comparative analysis of high school students in Kazakhstan. *外国語教育論集[Studies in Foreign Language Education]*, 46, 21-33.
https://tsukuba.repo.nii.ac.jp/record/2010027/files/SFLE_46-21.pdf

Appendix

SMAT Loudness Subtest Answer Sheet

Loudness Test: Is the second tone stronger or weaker than the first?										
Name										
Date					Hour					
School										
Grade										
Age				years				months		

Column										
	A		B		C		D		E	
1	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
2	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
3	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
4	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
5	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
6	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
7	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
8	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
9	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger
10	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger	weaker	stronger

SMAT Pitch Subtest Answer Sheet

PITCH Test: Is the second tone HIGHER or LOWER in pitch than the first?										
Name										
Date					Hour					
School										
Grade										
Age				years				months		

Column										
	A		B		C		D		E	
1	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
2	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
3	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
4	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
5	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
6	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
7	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
8	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
9	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
10	lower	higher	lower	higher	lower	higher	lower	higher	lower	higher

SMAT Timbre Subtest Answer Sheet

TIMBRE Test: Are the tones the SAME or DIFFERENT in Timbre (Tone Quality)?										
Name										
Date					Hour					
School										
Grade										
Age				years				months		

Column										
	A		B		C		D		E	
1	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
2	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
3	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
4	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
5	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
6	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
7	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
8	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
9	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.
10	same	diff.	same	diff.	same	diff.	same	diff.	same	diff.

SMAT Time Subtest Answer Sheet

TIME Test: Is the second tone LONGER or SHORTER than the first?										
Name										
Date					Hour					
School										
Grade										
Age				years				months		

Column										
	A		B		C		D		E	
1	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
2	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
3	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
4	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
5	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
6	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
7	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
8	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
9	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer
10	shorter	longer	shorter	longer	shorter	longer	shorter	longer	shorter	longer

SMAT Rhythm Subtest Answer Sheet

RHYTHM: Are the rhythm patterns of the two tone sequences the SAME or DIFFERENT?					
Name					
Date		Hour			
School					
Grade					
Age		years		months	

Column						
A		B		C		
1	same	diff.	same	diff.	same	diff.
2	same	diff.	same	diff.	same	diff.
3	same	diff.	same	diff.	same	diff.
4	same	diff.	same	diff.	same	diff.
5	same	diff.	same	diff.	same	diff.
6	same	diff.	same	diff.	same	diff.
7	same	diff.	same	diff.	same	diff.
8	same	diff.	same	diff.	same	diff.
9	same	diff.	same	diff.	same	diff.
10	same	diff.	same	diff.	same	diff.

SMAT Tonal Memory Subtest Answer Sheet

TONAL MEMORY: Which note (1st, 2nd, 3rd, and so on) differs in the pair of tone sequences?					
Name					
Date		Hour			
School					
Grade					
Age		years			

Column			
	A	B	C
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			