The Impact of Music Aptitude and L1 on Non-native Language Prosody Perception

Xinyi Pei

MSc in Applied Linguistics and Second Language Acquisition, 2025

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Xinyi Pei
Date:	18/08/2025

The Impact of Music Aptitude and L1 on Non-native Language Prosody Perception

Xinyi Pei

Department of Education

University of Oxford

St. John’s College

Word Count: 19936

Dissertation submitted in part-fulfilment of the requirements for the degree of

Master of Science in Applied Linguistics and Second Language Acquisition.

August 2025

Table of Contents

Acknowledgements

Music and language are two of the greatest passions in my life, and it still feels so surreal to have turned this personal interest into a completed academic work. I am truly fortunate to have spent a year delving into a subject that fascinates me so much.

I have played the piano for 17 years and always loved music, and my passion for language is what brought me to the ALSLA programme. Yet, I never imagined their connection could be the subject of my research. The idea took root during a supervision with my then-supervisor, Johannes, when he challenged me to purse a topic I was “truly interested in”. That simple question sparked a reflection on my life: I remembered watching a video that of a little girl with perfect-pitch mimicked Mandarin on her cello, and I realised how people I knew who spoke foreign languages with almost no accents also sang well and vice versa. That’s when it all connected -- music and language are indeed connected, so why not study this fascinating connection?

And so, here I am, under the support and guidance from so many people that I’m always grateful in my heart, I have finished this dissertation. My sincere gratitude goes to my supervisors. Johannes’s support throughout Michaelmas term and his humor made supervisions a joy and brought me great mental comfort, helping me find a clear path for my research. My deepest gratitude goes to Liz, who is, as Johannes righty said, “the smartest person I’ve ever seen!”. She’s always so organised, efficient, knowledgeable and professional, and could always spot potential problems with my research design, data analysis, or drafting the dissertation. She is always there to provide meticulous, to-the-point feedback, saving me from countless obstacles, and her encouragement when I encountered difficulties was a great comfort that kept me moving forward.

My gratitude also extends to all the teachers and professors during my journey at ALSLA. I’m especially thankful for Faidra, who was one of my interviewers and whose kind and lovely nature was a key reason I chose to join ALSLA, and whose course I have always enjoyed so much. Thanks for inviting me into her small research group, CreATE, where we discussed creative approaches like drama, singing, poetry, and debating. I always feel so grateful to be part of it, receiving both academic and mental support, and feeling a sense of belonging. A special thanks you to Cate, who gave me tremendous support, especially with stats. I will never forget the two-hour online meeting where she patiently guided me through R coding step-by-step, making the tedious task feel hopeful and possible. To all my other teachers, thank you for teaching me to do critical thinking and for leading me to the threshold of being an independent researcher.

Thank you to the ALSLA cohort, whose diverse experiences were a constant source of inspiration, and I am so grateful to have made some lifelong friends among you. Thank you to my friends in Oxford, Yu, Tong, Ella, Doris, Li, Nic, Sophia, April, Jesse, Hang..., whose companionship helped me fit in this completely new environment, whose individual shining points inspired and empowered me all the time, who walked beside me through the darkest nights and stood with me under the most splendid skies. Thank you to my old friends, Jade and Vivian, who, even across different time zones, were always there for me and lent me a shoulder to cry on. Thank you to my parents, who supported me both financially and mentally, and whose faith in my abilities and aspirations has been my source of strength and motivation.

Thank you, Oxford. Thank you for all the amazing views in this charming city, and the magnificent libraries, from St. John’s library, which gave me a cozy spot to meet the deadlines at midnight, to Radcliff Camera, which made my childhood dream of studying at “Hogwarts” came true, and to many more. Thank you for witnessing my struggles and growth. Thank you for bringing so many incredible people into my life and for filling my journey with memories that will continue to empower me for years to come.

List of Tables

List of Figures

Abstract

Mounting evidence suggests a relationship between music aptitude and speech prosody perception, particularly in second or foreign language contexts where perceptual demands are higher. However, existing studies often confound music training with innate aptitude, rely on binary musician/non-musician groupings, and rarely account for language background or L2 proficiency in nuanced ways. Addressing these gaps, this study investigates the intricate interplay among music aptitude, L1 background, English proficiency, and non-native English prosody perception in adult ESL/EFL learners. The study employed continuous measures of music aptitude (Mini-PROMS) and English proficiency (LexTALE), alongside a quantifiable L1 phonological distance metric, in a cross-sectional correlational design with 100 participants. Prosody perception was further measured through a multidimensional battery that isolates spectral and temporal cues.

Key findings indicate that overall music aptitude significantly and positively correlates with non-native prosody perception accuracy, even after controlling for L1 and language proficiency. Specific sub-components of music aptitude demonstrated distinct predictive power: musical tuning perception was a significant predictor for pitch-related prosody, while musical tempo perception significantly influenced timing-related prosody. English proficiency emerged as the most robust and consistent positive predictor across all analyses, highlighting its foundational role in second or foreign language development. Conversely, L1 background, whether treated as a continuous phonological distance or categorical rhythmic typology, did not show a statistically significant direct association or moderating effect on the music-prosody relationship within this sample.

These results align with the OPERA hypothesis. The domain-specific transfer observed between musical dimensions and prosodic features suggests a modularity in shared perceptual mechanisms. While the non-significant role of L1 background appears to challenge some existing models, it may reflect the mitigating influence of intermediate-to-high L2 proficiency within this sample. This research offers valuable theoretical insights into individual differences in L2/FL acquisition and suggests potential avenues for targeted pedagogical interventions focusing on holistic proficiency development and specific auditory training.

I. Introduction

Human communication takes many forms, but few are as richly structured and universally present as language and music. Though traditionally studied as separate domains—one rooted in linguistics, the other in the arts—an increasing body of research has begun to uncover deep cognitive and neural connections between the two. As Patel (2007) observes, music and language are best understood not as monolithic systems but as “constellations of subprocesses” (Patel, 2007, p. 398), some of which overlap and interact in complex ways. Investigating these interactions, in a broad sense, may offer unique insights into how the human mind transforms sounds into structured, meaningful experience. In a narrower sense, from the perspective of language acquisition, this may explain an important component of individual differences: why, even when controlling for age, input, and many other factors, adult learners still acquire a second language (L2) or foreign language (FL) differently, with some attaining near-native levels while others having strong foreign accents or frequent grammar errors. This dissertation thus adopts this interdisciplinary perspective, exploring the interplay between music aptitude and language perception. Music aptitude, defined as the measure of an individual's potential to learn music (Gordon, 1989), is considered a relatively stable capacity and is largely independent of acquired music training experiences or musical knowledge, and is the target of this study.

Among the components of language, prosody, “the melodic and temporal properties of speech that form the suprasegmental components of the phonology of a language” (Jansen et al., 2023, p. 1), is considered to be especially connected to music, as both are composed of acoustic cues, have similar structures, and share similar communicative functions like expressing emotions (Heffner & Slevc, 2015). Empirical studies have found that music training positively affects prosody perception (e.g., Marques et al., 2007; Schön et al., 2004), and that music perception and prosody perception abilities are positively correlated as reported in studies using either behavioral methods or electrophysiological methods (Correia et al., 2022; Marques et al., 2007; Moreno et al., 2009; Perrachione et al., 2013; Schön et al., 2004; Wong et al., 2007). Interestingly, this correlation is especially salient in non-native language (i.e., L2 or FL—the following sections will use non-native language as an umbrella term) prosody perceptions, which may be a result of a ceiling effect for native language (Jansen et al., 2023).

While such music-language interaction is grounded in frameworks like the OPERA hypothesis (Patel, 2011), which posits that music training can enhance speech processing because both domains rely on shared, acoustically precise mechanisms, this primarily focuses on first language (L1) acquisition and processing, and its explanatory power remains constrained in cross-linguistic contexts. Specifically, its premise of shared acoustic precision requirements fails to fully account for the complexities of non-native language acquisition, where learners often experience cross-linguistic phonological interference that the hypothesis does not directly address. This theoretical limitation is reflected in the empirical literature, where significant variability in findings persists despite meta-analyses reporting an overall moderate positive relationship (e.g., Thompson et al., 2025).

Additionally, previous studies are often limited by methodological shortcomings. Many studies, for instance, roughly categorise participants into a binary distinction between musicians and non-musicians, an oversimplification that fails to differentiate between innate music aptitude and the effects of acquired music training, necessitating the need to directly measure musical aptitude as a continuous spectrum. Other potential confounding factors have also been identified, including the learner’s L1 background and their proficiency level in the target language.

Addressing these gaps, this study seeks to understand how distinct aspects of music aptitude—melody, tuning, rhythm, and tempo perception—predict the ability to perceive prosodic features like pitch and timing in English as a non-native language. Furthermore, the study explores the influence of L1 background (measured as a numerical “phonological distance”) and English proficiency on this relationship. By incorporating continuous measures of music aptitude, language proficiency, and phonological distance, the research aims to investigate the interaction between music aptitude and non-native language prosody perception in a more rigorous and nuanced way.

II. Literature Review

This chapter provides a comprehensive review of the theoretical and empirical literature that underpins the relationship between music aptitude and non-native (L2 or FL) language prosody perception. The review begins by examining the broader evidence for shared neural resources and mechanisms behind music and language processing (Section 2.1). It will then narrow to the specific role of prosody in non-native language perception, defining its key features and examining the unique challenges faced by non-native learners (Section 2.2). Following this, a comprehensive review of empirical studies on the connection between music ability and prosody perception will be presented. This review will specifically differentiate between findings based on musicianship status, formal music training interventions, and, most critically, innate music aptitude (Section 2.3). The chapter will then address potential moderating factors, including the influence of L1 background and L2 proficiency level (Section 2.4), and finally highlighting the gaps in the literature that this dissertation aims to address (Section 2.5).

2.1 The Relation Between Music and Language

Music and language are unique to human beings and share many similarities (Patel, 2007). Both domains rely on a limited set of acoustic features, such as pitch, rhythm, loudness, and timbre. When perceiving these acoustic cues, the brain categorises them into discrete representations, or sound categories, which are then hierarchically structured to convey complex and meaningful information. For instance, phonemes are structured into words and sentences, while musical notes are structured into melodies and larger compositions (Patel, 2007; Schön & Morillon, 2018).

This structural and functional parallelism between music and language suggests that the two domains may engage shared perceptual and cognitive processes. Therefore, researchers have been exploring whether and how music and language perception and processing rely on shared neural resources and mechanisms, and whether and how these overlaps lead to cross-domain transfer, which will be briefly reviewed in the following sections.

2.1.1 Shared Neural Resources Between Music and Language Processing

Studies have demonstrated that music and language processing, especially syntactic processing, involve overlapping brain regions and share neural resources.

To begin with, neuroimaging research has consistently shown that music and language syntactic processing rely on overlapping brain regions. For instance, studies using magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) have located brain activity for musical harmonic processing in areas conventionally known to be associated with language, such as Broca's and Wernicke's areas (Maess et al., 2001; Koelsch et al., 2002). Further supporting this, Patel et al. (1998) found that musicians produced a similar brain response (P600 ERP) when processing both syntactically incorrect sentences and musically incongruous chord sequences. This suggests that the brain uses a common mechanism to detect structural violations in both domains, highlighting a shared neural foundation for music and language.

These evidence have led Patel to propose the Shared Syntactic Integration Resource Hypothesis (SSIRH), which posits that music and language share a common set of neural processing resources specifically for syntactic integration. This hypothesis not only integrates previous empirical findings but also effectively accounts for seemingly contradictory evidence from neuropsychology, such as cases of aphasia without amusia (language impairment with intact musical ability). According to SSIRH, such dissociations are explained not by completely separate integration resources, but rather by damage to the domain-specific regions in the temporal lobes that store long-term knowledge of linguistic or musical rules, while the shared, general-purpose processing regions (responsible for integrating information) remain intact (Patel, 2003). Crucially, the SSIRH further predicts that if these shared syntactic integration resources themselves are impaired, then both linguistic and musical structural processing should be affected. This is supported by a later study by Patel et al. (2008), who found that aphasia patients who suffered from impaired syntactic language processing also showed a significant deficit in processing musical structure. This was demonstrated in a harmonic priming task, where aphasia patients, unlike healthy individuals, did not respond faster to harmonically related chords, indicating that their ability to process hierarchical musical structure was impaired. This finding, where a deficit in one domain's structural processing co-occurs with a deficit in the other, strongly points to shared syntactic integration resources between language and music.

Furthermore, a key prediction of the SSIRH is that because these resources are limited, syntactic processing in music and language should interfere with each other when performed concurrently. Addressing this prediction, Koelsch et al. (2005) used ERP to show that musical syntactic violations (e.g., out-of-key chords) attenuated a typical linguistic syntactic violation marker (left anterior negativity/LAN) when both occurred simultaneously. Similarly, behavioral studies by Fedorenko et al. (2009) and Slevc et al. (2009) demonstrated that out-of-key chords disrupted syntactic processing in language tasks, an effect not observed with purely acoustic musical irregularities (e.g., louder sounds). Collectively, these findings provide strong evidence for the SSIRH, highlighting that music and language share limited syntactic integration resources.

However, one limitation is that these paradigms often rely on highly controlled dual-task designs that may not reflect natural processing of musical or language input. Moreover, most studies focused on Western music and native speakers of Indo-European languages, limiting generalisability. Nevertheless, these findings collectively support the SSIRH, highlighting the shared resources between music and language processing.

2.1.2 Shared Mechanisms Behind Music and Language Processing

Another similarity between music and language, and also one that marks music and language different from other sound systems and unique to human beings is that, the sound categories in music and language are not physical entities like a chemical substance, but are mental constructs shaped by a learned framework of auditory categories (Patel, 2007). This is because, human beings perceive musical and linguistic sounds through the mechanism of categorical perception (CP), where the brain does not perceive acoustic cues on a continuous spectrum, but instead places them into discrete perceptual categories. For example, in language, the continuous change in Voice Onset Time (VOT) that distinguishes a /b/ from a /p/ is heard as a sudden switch from one sound to the other at a specific boundary. Differences within a category are effectively ignored, while differences across the category boundary are highly salient. Similarly, in music, the continuous spectrum of pitch frequencies is organised into discrete notes (e.g., A, B, C). A note that is slightly sharp or flat is not heard as a new note, but as an out-of-tune version of a familiar note (Weidema et al., 2016; Yashaswini & Maruthy, 2020)

However, these categorical boundaries are not universal across all languages or musical cultures. For instance, the sounds /l/ and /r/ are distinct phonemes in most languages, but are often considered allophones of a single phoneme in Japanese. This means that the acoustic differences between /l/ and /r/ fall within the same perceptual category for native Japanese speakers, who have learned to treat them as variations of a single phoneme and find it hard tell the two sounds apart. Similarly, in music, the difference between major and minor thirds—important melodic intervals in Western European music—may not hold the same significance in some New Guinean cultures, to whom both are perceived as variations within a single sound category (Chenoweth, 1980). These examples demonstrate that the human auditory system is highly plastic, with its perceptual categories being shaped by learned experiences and context, where language and music backgrounds may have significant impact on the categorical perception of linguistic and musical sounds. This plasticity and the shared underlying mechanisms discussed above have led studies to investigate whether skills and knowledge gained in one domain can "transfer" to the other, a phenomenon known as cross-domain transfer.

2.1.3 Cross-domain Transfer

To explain this transfer, Patel (2011) proposed the OPERA Hypothesis, a theoretical framework that outlines the conditions under which musical training can benefit language processing. OPERA is an acronym for five key criteria: the neural processing involved in music training must be **O**verlapping with language, require **P**recision for fine-grained acoustic distinctions, involve **E**motion, require **R**epetition to automate learning, and rely on focused **A**ttention. According to this hypothesis, music training enhances auditory processing skills that meet these criteria, leading to a beneficial transfer to language skills.

Empirical evidence from training interventions provides strong support for a causal cross-domain transfer. For example, musical training specifically designed to improve rhythmic perception and production resulted in improved phonological and reading skills

(Bhide et al., 2013; Cogo-Moreira et al., 2013; Flaugnacco et al., 2015). Beyond behavioral improvements, research also indicates that musical training can sharpen low-level auditory processing in subcortical brain regions, which enhances the neural encoding of linguistic pitch patterns (Wong et al., 2007).

Another significant area of overlap facilitating cross-domain transfer lies in the expression and perception of emotion. Both music and language convey emotions like happiness, sadness, anger, and fear through a shared set of acoustic cues, including tempo, intensity, and frequency patterns (Juslin & Laukka, 2003), and research has shown that musical training can improve the ability to perceive emotional prosody in speech (Thompson et al., 2004). These findings collectively reinforce the idea that music and language processing share similar mechanisms and resources, leading to positive transfer between domains.

2.1.4 Summary

Bringing it together, music and language share structural parallels, rely on overlapping resources and similar mechanisms, and the plasticity of the human auditory system allows the cross-domain transfer from music to language. This leads us to a more detailed examination of how certain aspects of language are connected to music, where prosody, the "music" of language and a key component of non-native language perception, comes into focus.

2.2 Prosody and Its Relevance to Non-native Language Perception

2.2.1 What is Prosody and Why It is Important

In language, prosody is defined as "the melodic and temporal properties of speech that form the suprasegmental components of the phonology of a language" (Jansen et al., 2023, p. 1). "Suprasegmental" refers to the aspect of speech that is hierarchically above segments, from the basic unit of the syllable to words, phrases, clauses, and sentences (Peppé, 1998). Prosody can be understood as the musical aspects of speech, with the melody-related properties like intonation and rhythm-related properties like stress (Kolinsky et al., 2009). They are composed of acoustic features like pitch, accents, pauses, and loudness, and are acoustically reflected in fundamental frequency, duration, and intensity (Lehiste, 1970; Speer

& Blodgett, 2006; Wagner & Watson, 2010).

Functionally speaking, prosody conveys linguistic and emotional information beyond individual words. Through its suprasegmental features, prosody helps structure utterances by grouping words together (also called prosodic phrasing) and marking syntactic/semantic relationships, signals speech focus and prominence through emphasis, indicates speech acts and discourse functions (e.g., questions vs. statements, turn-taking cues), and expresses paralinguistic information like speaker attitude and emotion (Wagner & Watson, 2010).

Prosody is an important aspect of language acquisition because it is crucial to language comprehension. For infants, their acquisition of L1 is solely auditory, and they tend to show a preference for the prosody of their native language as early as three days of age (Mehler et al., 1988). For adults, a large proportion of their language input is spoken language, and many dialects have no written form (Speer & Blodgett, 2006). To understand a speaker's intended meaning, the listener must rely on prosodic cues to identify the boundary of syllables, words, syntactic constituents, and the pragmatic status of items in the discourse, especially when multiple meanings can arise from the same set of phoneme strings (Speer & Blodgett, 2006). A typical example is the utterance sets "What's that ahead in the road?" and "What's that, a HEAD in the road?", where the phoneme strings are identical, but different prosodic cues like pauses or pitch patterns can lead to different meanings. Supporting this, empirical studies have found that prosodic cues can clarify ambiguous syntactic structures and understand complex sentences (e.g., Frazier et al., 2006; Jasmin et al., 2018; Langus et al., 2012). Furthermore, prosody can even modify our understanding of an utterance, as Hirst (2004) noted, "when there is a discrepancy between the prosody of the utterance and its overt semantic content, we usually trust the prosody rather than the semantics" (2004, p. 1), which is typically seen in how modifying the intonation of an overtly genuine compliment turns it into a sarcastic criticism, which further confirms prosody's essential role to language comprehension.

2.2.2 Two Major Aspects of English Prosody: Phrase Boundary and Prosodic Prominence

The effect of English prosody on language comprehension is primarily realised through two core features: (1) *boundaries*, which segment speech into meaningful chunks like phrases or words, and (2) *prominence* (stress, accent, or focus), which highlights key elements within these chunks (Cole, 2015). Both features are characterised by prosodic cues varying across three dimensions: duration, fundamental frequency, and intensity. Duration refers to time-related cues including length of speech and pauses, and is identified as the most reliable cue in disambiguating syntactic structures (Lehiste, 1973). Fundamental frequency (F0) is the lowest frequency of a sound wave produced by vocal cord vibrations during speech. It determines the perceived pitch of a voice, where a higher F0 is perceived as a higher pitch (Menn, 1982). Intensity refers to the loudness of a sound wave, but is less studied in prosodic analysis and its relevance to prosodic boundaries is inconsistent across speakers (Kim et al., 2006), thus is excluded from this dissertation's scope.

This theoretical categorisation of prosodic features has methodological implications for both the measurement of prosodic perception and the design of experimental stimuli. For example, this framework is operationalised in a contemporary assessment tool—the Multidimensional Battery of Prosody Perception (MBOPP; Jasmin et al., 2021), which partitions prosody perception into two subtests: (1) Linguistic Focus, which evaluates sensitivity to prominence (i.e., sentence stress) of target words, and (2) Phrase Boundaries, which assesses the ability to detect phrase boundaries in a compound sentence. In both subtests, the stimuli are meticulously manipulated to isolate the effects of either pitch or timing cues. In the pitch-only condition, two versions of the same spoken text differ only in their pitch contour, while in the timing-only condition, the two versions differ only in duration. This bipartite test structure and its unique manipulation of conditions not only reflect the core dichotomy of prosodic function (Cole, 2015) but also addresses a critical gap in earlier instruments (e.g., PEPS-C; Peppé & McCann, 2003). By precisely disassociating pitch cues and duration cues, the MBOPP's design allows for granular insights into the specific functional contributions of each type of prosodic cue to prosody perception.

2.2.3 Prosody Perception in Non-native Languages

Though languages are rich with prosodic cues, understanding sentences in a non-native language that rely on prosodic cues can be effortful and difficult for L2/FL learners.

Understanding this difficulty begins with the fundamental mechanism of prosody perception, which, like individual phonemes, operates through Categorical Perception (CP) (Remijsen & van Heuven, 1999; Schneider & Lintfert, 2003). The continuous variations in pitch, duration, and intensity that define prosody are thus not perceived as a continuum, but are categorised into meaningful units, such as a question intonation (rising pitch) or a statement intonation (falling pitch) (Schneider et al., 2009). As mentioned before, the human auditory system is highly plastic, and is "trained" by one's L1 to perceive the L1-specific categorical boundaries. More specifically, when a person encounters a new sound in a non-native language, they tend to assimilate it to the closest existing sound category in their L1. This phenomenon has been summarised in Best et al.'s (1994) Perceptual Assimilation Model (PAM), which builds on CP to explain how native language experience acts as a filter for non-native sounds. For example, French speakers, whose language has fixed stress, when hearing two English words with distinct lexical stress (e.g., the noun and verb form of "conduct"), would assimilate English stress contrasts into their existing L1 categories, making it difficult to perceive them as different, as supported by empirical evidence from studies like Dupoux et al. (1997).

Furthermore, the Automatic Selective Perception (ASP) Model (Strange, 2011) adds an important layer to the PAM by focusing on the role of attention and cognitive resources. It posits that native listeners develop highly efficient and automatic selective perception routines, where their brains automatically attend to the most reliable acoustic cues that distinguish sounds in their L1 and ignore irrelevant ones. When an L2 learner encounters a new language, however, their L1's automatic routines are no longer effective, and they may not naturally attend to the relevant prosodic cues in the new language. This requires the learner to consciously employ greater cognitive resources and attention to perceive and process these new patterns, making the process slow and effortful. This is supported by empirical evidence like Asano's (2018), which found that German learners of Japanese showed significantly increased reaction times and error rates in discriminating Japanese consonant length contrasts under heightened task demands (e.g., increased memory load). Notably, these effects were much less pronounced for native Japanese listeners, highlighting the additional cognitive load faced by non-native listeners.

Together, these models explain the challenges faced by L2/FL learners during non-native prosody perception: Firstly, as described by the PAM, learners' brains tend to assimilate non-native prosodic cues to existing L1 categories, hindering their ability to perceive differences in the non-native prosody that do not exist in their L1. The ASP model further explains that learners fail to automatically attend to the specific prosodic cues (e.g., a specific pitch rise, a subtle lengthening of a syllable) that are critical for marking meaning in the target language, and the activated L1 prosodic cues may be irrelevant or even misleading in the target language, potentially leading to negative transfer effects. Therefore, successful non-native prosody perception frequently requires additional cognitive attention and resources for processing acoustic features.

Given these challenges, and based on the OPERA hypothesis (Patel, 2011), musical ability is theoretically positioned to potentially facilitate non-native prosody perception. Specifically, the precise processing skills for acoustic features like pitch and rhythmic patterns honed or inherently strong in music perception may positively transfer to, and potentially compensate for, the difficulties in perceiving non-native prosody. This suggests that music ability may help learners better detect and interpret the unfamiliar prosodic cues necessary for L2/FL comprehension.

2.3 Music Aptitude and Its Influence on Non-native Prosody Perception

Having reviewed the importance of prosody to language comprehension and its relevance to music, it is not surprising that many studies have explored the influence of music abilities on prosody perception. Although a tentative conclusion of a positive relationship can be reached, the various methodologies used, the incongruent results and methodological limitations across existing studies necessitate further exploration into this topic.

2.3.1 Defining Key Terms: Music Aptitude, Music Training, and Music Ability

The concept of music(al) ability can be understood from a variety of definitions and scopes. Some define it as an innate giftedness or natural talent (Gagne, 1999; Law & Zentner, 2012) that is independent of training, while others define it as the result of deliberate practice or environmental and intrapersonal catalysts (Sloboda & Howe, 1991). It is also interchangeably used with other terms like "competence" (Wallentin et al., 2010) or "musicality" (Müllensiefen et al., 2014). This dissertation adopts Choi's (2022) working definition of music ability as "the perceptual sensitivity to music-related acoustic features such as pitch and rhythm" (2022, p. 10) and interprets it as the collaborative result of genetic predispositions and environmental factors (Kragness et al., 2020). The genetic determining factor is music aptitude—an individual's potential to learn music (Gordon, 1989), which is innate and is independent of acquired training or musical knowledge. The environmental factor is the acquired music training experience, which might influence the perceived music abilities, though the exact influence needs further investigation.

The relationship between musical training and prosody percepetion has been explored mainly in two ways: (1) cross-sectional studies that explore the correlations between musicianship (i.e., whether participants are "musicians" based on prior experience, often defined by 6-7+ years of music training) (Choi, 2022) and prosody perception, and (2) intervention studies that provide participants with actual music training during the study (e.g., Moreno et al., 2009; Nan et al., 2018), ranging from 6-minute learning tasks (Ong et al., 2016) to 1-year lessons (Thompson et al., 2004) and compare their performance before and after the intervention. The following sections will address these different types of studies accordingly.

2.3.2 Studies on Musicianship

Cross-sectional studies on musicianship consistently report that musically experienced adults are better at detecting and identifying patterns or deviations in non-native prosody compared to their non-musical counterparts, suggesting a facilitative effect of musicianship on non-native prosody perception. This "musician advantage" encompasses a wide range of prosodic features, including enhanced sensitivity to acoustic cues like F0 variation and syllable duration (Marie et al., 2011; Marques et al., 2007; Sadakata & Sekiyama, 2011), improved identification and discrimination of lexical tones (Alexander et al., 2005; Bradley, 2012, 2018; Choi, 2020; Delogu et al., 2010; Götz, 2023; Lee & Hung, 2008; Maggu et al., 2018; Marrie, Delogu et al., 2011; Weidema et al., 2016; Wong & Perrachione, 2007) and lexical stress (Boll-Avetisyan et al., 2016; Kolinsky et al., 2009), perception of sentence-level intonation contours (i.e., question/statement intonation, as in Martínez-Castilla et al., 2019; or pitch change at the end of the sentence creating either congruous or incongruous prosody with the preceding context, as in Deguchi et al., 2012 and Toh et al., 2023) and sentence metrics (i.e., rhythmic structure and syllabic duration, as in Marrie, Magne et al., 2011), identification of sentence prosodic focus or phrasing (Martínez-Castilla et al., 2019; Obergfell et al., 2021; Stepanov et al., 2018), and identification of emotion through prosody (Fuller et al., 2014; Thompson et al., 2003).

Despite these consistent findings, these studies suffer from two major methodological limitations. Firstly, their experience-based nature inherently prevents them from disentangling the effects of acquired music training from innate music aptitude. Any observed group differences cannot be solely attributed to training, as individuals with high innate musical talent might self-select into musical training, confounding the causal interpretation. Secondly, these studies frequently oversimplify musicianship into a binary "musician/non-musician" distinction, overlooking its inherent heterogeneity, as musicianship encompasses a wide array of experiences, including various lengths and types of training (e.g., type of instruments played).

For example, Kolinsky et al. (2009) compared French musicians and non-musicians on English stress sequence recall, finding musicians outperforming non-musicians. However, this conclusion is problematic because participant grouping was based solely on self-reported musical experience without assessing innate aptitude. Therefore, the observed advantage could stem from natural musical talent rather than training. This is especially a concern as most of the "musicians" in this study were absolute pitch possessors, which is a rare and professional musical talent, suggesting that their advantage might be rooted in this innate perceptual talent. Furthermore, the majority of the musicians in this study played the violin, while the others play the piano or other instruments. introducing another layer of heterogeneity. Choi et al. (2024) later demonstrated that pitched instruments (e.g., violin) musicians exhibit a unique advantage in lexical tone perception compared to unpitched (e.g., piano) musicians and non-musicians, underscoring that different musical experiences can yield different music-to-language transfer effects. This highlights the methodological limitation of simplistic binary groupings of musicianship, as musicians constitute a heterogeneous group with significant individual differences in both experience and aptitude.

A later study by Delogu et al. (2010) possesses a methodological advancement, as they grouped participants not only by musical training experiences, but also by their music aptitude through a standardised test of musical intelligence (Wing, 1968). Though they found a positive relationship between greater melodic ability and enhanced lexical-tonal discrimination performance, they did not report how musicianship specifically correlated with their melodic perception ability. Moreover, the separate comparisons between musicians/non-musicians and between high, medium, and low melodic perception groups made it hard to draw unified conclusions, thus undermining the general claim that "musicians outperform non-musicians" and obscuring the exact role of music training. These ambiguities in musicianship studies thus underscore the need for a more direct examination of causality, which necessitates turning our attention to studies employing music training interventions to draw more sound conclusions on the role of music training.

2.3.3 Studies on Music Training Through Interventions

Studies employing controlled music training interventions are relatively few and present a mixed picture. Most targeted L1 prosody perception, including Mandarin lexical tone discrimination in children (Nan et al., 2018) and pitch incongruity perception in sentence intonation contour among Portuguese children (Moreno et al., 2009), and also two focusing on adults' emotional prosody perception in Hebrew (Bodner et al., 2012; Mualem & Lavidor, 2015). Despite reporting positive results, the significant variety in training approaches (e.g., 30-minute lessons on recognising emotions in music, as in Mualem & Lavidor, 2015; vs. musical instrument lesson as in Nan et al., 2018; vs. group lessons on specific musical dimensions, as in Moreno et al., 2009), coupled with the heterogeneity of prosody perception types, target languages, and age groups included, makes it hard to draw any general conclusion.

In terms of adults' non-native prosody perception, studies are even fewer. Among the few existing studies, Ong et al. (2016) provided unprecedented evidence that even ultra-short musical exposure (6 minutes of distributional learning with novel musical chords) could enhance linguistic pitch perception. By exposing L1-English speakers to either a Thai lexical tone minimal pair or a novel musical chord minimal pair and testing their post-training discrimination in both domains, they found that performance improved both within-domain (e.g., musical chords training to chord discrimination) and cross-domain (e.g., musical chords training to Thai tone discrimination) minimal pairs, suggesting enhanced processing of pitch cues generalised across domains. This finding aligned with Patel's (2011) OPERA hypothesis by suggesting a positive transfer between music and language. However, a key limitation of this study was its use of oversimplified stimuli—isolated lexical tones and musical chords—which lacked domain-specific characteristics (e.g., sentence-level intonation or melodic context), raising concerns that participants may have treated them as general acoustic cues rather than distinct linguistic or musical stimuli.

Building on Ong et al.'s (2016) findings, Barbaroux et al. (2021) extended the intervention duration to 3 hours, demonstrating that sustained pitch discrimination training (with harmonic complex sounds) significantly enhanced non-native Thai lexical tone discrimination in L1-French non-musicians, with no improvement in an intensity (loudness) control group. This study strengthened the causal relationship of music-to-language transfer by allowing for more robust learning and ruling out general auditory attention effects with the intensity control group. However, both Ong et al. (2016) and Barbaroux et al. (2021) focused exclusively on pitch, neglecting other prosodic dimensions like duration, which is critical in languages with distinct timing patterns (e.g., stress-timed vs. syllable-timed languages). Additionally, neither of these studies measured baseline musical aptitude, leaving open the possibility that pre-existing pitch sensitivity mediated improvements, a confound later highlighted by Wiener & Bradley (2023).

Wiener and Bradley (2023) introduced an 8-week intervention that either provided Mandarin classroom training (+L2), music training (+Music), or combined training (+L2+Music), finding that the +L2+Music group showed significantly greater improvement in Mandarin tone sensitivity than the other groups, while the +Music group showed only marginal, non-significant improvement, contrasting with Barbaroux et al.'s (2021) findings. This might be attributed to large individual differences in baseline pitch sensitivity among participants, as reflected from the pre-test, suggesting that participants with higher pre-intervention pitch sensitivity might have benefited more, while the less sensitive participants pulled down the overall group mean. This highlights a general limitation of intervention studies: the majority failed to adequately control for or separate the effect of baseline individual differences in innate perception abilities or music aptitude, which may confound the observed training effects. Additionally, Wiener & Bradley (2023) also acknowledged a fundamental challenge for intervention studies--while they are crucial for exploring causal effects of music training, their often short-term nature (e.g, two months in this study) may be insufficient to produce the robust, long-term changes observed in highly trained musicians, thus presenting a paradox: to isolate the causal effect of music training, interventions are required; yet for those effects to manifest significantly, long-term exposure akin to comprehensive musicianship is often necessary, which typical intervention designs are unable to provide. These hard-to-solve methodological challenges turn our attention to another line of research that could offer greater methodological rigor--studies exploring the fundamental relationship between innate music aptitude and prosody perception.

2.3.4 Studies on Music Aptitude

Compared to the complexities of measuring music training effects, assessing innate music aptitude offers a more reliable approach, as it is inherently stable over time and less susceptible to the confounding variables associated with varied musical experiences. Therefore, a rigorous exploration of the music aptitude-prosody perception relationship necessitates the use of a psychometrically sound test that can minimise the influence of music experiences.

However, most available test batteries (summarised in table 2.1) primarily assess perceived music abilities, which often reflect the mixed effects of music experiences and music aptitude, thereby conflating innate aptitude with learned skills. This calls the need to further confirm the aim of tests and select the ones that most effectively reflect innate music aptitude.

Table 2.1

Summary of Popular Test Measures of Music Aptitude

Test Measure	Proposer & Year	Applied by (selected examples)
Advanced Measures of Music Audiation (AMMA)	Gordon, 1989	Bowles et al., 2016; Cooper & Wang, 2012; Kempe et al., 2015; Shockley, 2022
Gordon Musical Aptitude Profile	Gordon, 1965	Cooper, 2010
Montreal Battery of Evaluation of Amusia (MBEA)	Peretz et al., 2003	Chen et al., 2016; Dittinger, 2016; Morril et al., 2015; Nan et al., 2010; Qin et al., 2021; Wang & Peng, 2014; Zhang et al., 2018
Montreal Battery of Evaluation of Musical Abilities (MBEMA)	Peretz et al., 2013	Cui & Kuang, 2019
Musical Ear Test (MET)	Wallentin, 2010	Chen et al., 2016; Choi, 2022
Wing's Musical Aptitude Test	Wing, 1968	Bowles et al., 2016; Slevc & Miyake, 2006
Goldsmiths Musical Sophistication Index (Gold-MSI)	Müllensiefen et al., 2014	Correia et al., 2022
Seashore's Test of Musical Talents	Seashore, 1919	Faßhauer et al., 2015
Researcher-developed pitch or timing perception tasks	E.g., Musacchia et al., 2008	E.g., Musacchia et al., 2008
Profile of Music Perception Skills (PROMS)	Law & Zentner, 2012	Li et al., 2024

Among the available test materials, only a few explicitly claimed they measured specifically the innate music aptitude. The earlier instruments, including Seashore's measures (1919) and Wing's Musical Aptitude Test (1968), are unsuitable for modern research due to a lack of modern psychometric validation, as they suffer from stimulus imbalance (e.g., inconsistent timing/timbre in human-performed recordings), outdated audio formats, and unreported discriminant/convergent validity (Zentner & Strauss, 2017), and were therefore not considered for test selection.

More recent and psychometrically valid tests include Gordon's AMMA. It was built on Gordon's theoretical framework that distinguishes music aptitude as "a student's potential to achieve in music" (1989, p.4), which stablises around age nine and after which is not expected to change regardless of further experience, from music achievement, which is "what a student has already learned" (1989, p. 4). Building on this framework, AMMA aims to reveal the concealed music aptitude, regardless of any received music instruction (Gordon, 1989). However, AMMA's scope is relatively narrow, focusing only on tonal and rhythmic aptitude while neglecting other crucial dimensions like tempo and harmony (Zentner & Strauss, 2017). It also exhibits a cultural bias, as its stimuli rely predominantly on Western tonal and rhythmic structures, disadvantaging individuals from non-Western musical traditions (Brown & Jordania, 2013).

Another widely used test battery is the MBEA (Peretz et al., 2003). It primarily served as a diagnostic tool for amusia--the lifelong inability to process basic musical skills despite normal intelligence and memory, and thus can be regarded as a tool to assess the universal and innate capacities of musical perception. However, empirical studies using the MBEA have revealed that musicians achieve significantly higher scores across all indices compared to non-musicians, and scores of musicians have been found to correlate with their years of musical practice (Houghton, 2021). This challenges MBEA's construct validity as a measure of purely innate abilities, suggesting that it also captures learned musical skills. Additionally, MBEA exhibits a ceiling effect for neurotypical individuals, as reported in Qin et al.'s (2021) study, where 75% of participants scored above 90% and 7 participants (22%) achieved perfect scores of 100%. This invalidates its use for assessing individual differences in normal-range musical abilities (Peretz et al., 2003).

In contrast, the PROMS (Law & Zentner, 2012) stands out as a more psychometrically robust and suitable battery for measuring innate music aptitude. It was specifically developed to overcome the limitations of earlier assessments that relied on musicianship status (Law & Zentner, 2012). To minimise the influence of acquired musical knowledge and cultural biases, PROMS focuses on the elementary aspects of music which can be found across musical systems and traditions, and deliberately used "basic and abstract rather than complex and contextualised musical stimuli" (Law & Zentner, 2012, p. 4). By reducing stimuli to their most fundamental perceptual elements (e.g., pure tones, simplified rhythmic patterns, neutral timbres), PROMS effectively aims to bypass learned musical experiences and cultural knowledge, thereby reflecting more universal and potentially innate musical perceptual sensitivities.

Moreover, PROMS is designed to measure specific musical perceptual skills via separate subtests, including nine distinct dimensions in the full version, and four dimensions in the concise version, the Mini-PROMS (melody, accent, tempo, and tuning). Specific design features were also incorporated into each subtest to minimise the impact of prior musical experience. For example, in the Melody subtest, melodies are monophonic with constant rhythms, and a harpsichord timbre is chosen for its neutrality and unfamiliarity compared to common instruments, reducing advantage for musically trained individuals. Similarly, in the Tuning subtest, a simple C chord was used due to its culture-free nature. For the Accent and Tempo subtest, stimuli were chosen from various timbres (e.g., drums, bass, harmony) to minimise bias from instrument preferences due to musical experiences. All these contribute to PROMS' validity to minimise the influence of music experiences and reflect innate music aptitude. This validity is further confirmed by reported correlations with other established aptitude tests like Gordon's AMMA.

Additionally, PROMS and its various shortened versions all demonstrate satisfactory psychometric properties, exhibiting high internal consistency and test-retest reliability (Law & Zentner, 2012; Zentner & Strauss, 2017). Moreover, PROMS has shown a very low correlation with backward digit span, a standard measure of working memory (Kunert et al., 2016), suggesting that its results do not reflect general cognitive abilities like working memory, which may otherwise confound the measurement of pure music aptitude and is a limitation seen in many other batteries like the MET (Wallentin, 2010).

Taken together, the psychometric soundness, robust construct validity, and careful design to minimise the influence of acquired music experience make PROMS and its shortened version ideal for assessing individual differences in innate music aptitude. Its use in the current study thus represents a methodological advancement from previous studies and contribute to a clearer understanding of the music-prosody relationship.

2.4 Potential Moderators: The Role of L1 and Language Proficiency

2.4.1 The Role of L1

As previously mentioned, the Perceptual Assimilation Model (PAM; Best et al., 1994) suggests that the human auditory system is highly plastic, with L1 acting as a filter for non-native sounds, leading to the assimilation of upcoming new sounds to the closest existing sound categories in the L1. This theoretically hypothesises that different L1s may influence how non-native sounds are perceived, which is supported by empirical evidence on the differences between L1 toned and non-toned language speakers' pitch perception (e.g., Alexander et al., 2005; Chan & Leung, 2020). Specifically, the phonological distance between a learners' L1 and the target non-native language could, in theory, directly influence non-native sound perception, with smaller L1-L2 distances predicting better L2 perception.

Additionally, studies also suggest that L1 background may moderate the cross-domain transfer effect between music and language. For instance, research highlights significant differences in absolute pitch (AP) acquisition between tonal and non-tonal L1 speakers. Deutsch et al., 2006 found that 60% of Mandarin-speaking musicians who began training between ages 4-5 demonstrated AP, compared to only 14% of English-speaking musicians with similar training backgrounds. This suggests that the categorical perception of lexical tones in Mandarin may have enhanced musical pitch categorisation, which is further supported by neuroimaging studies showing overlapping neural substrates for linguistic and musical pitch processing in tone language speakers (Wong et al., 2007). Similarly, Cooper and Wang (2012) found that native speakers of tone languages (e.g., Thai, Mandarin, Cantonese) show enhanced sensitivity to pitch contrasts in both linguistic and musical domains. This striking connection between L1 tonal characteristics and both linguistic and musical pitch perception abilities theoretically positions L1 as a potential moderator of the relationship between music perceptual ability and non-native prosody perception.

However, L1 differences extend beyond the above-mentioned tonal versus non-tonal distinction. Languages also differ in rhythmic types, commonly catogorised as stress-timed, syllable-timed, and mixed (mora-timed) (Abercrombie, 1967; Dauer, 1983; Grabe & Low, 2002; Nespor et al., 2011; Ramus et al., 1999). Stress-timed languages (e.g., English, German, Dutch) feature near-equal intervals between stresses or rhythmic feet (i.e., the time between stressed syllables is roughly equal), whereas syllable-timed languages (e.g., Mandarin Chinese, French, Korean) exhibit successive syllables of near-equal length (Grabe & Low, 2002). A third type, mora-timed languages (also referred as “mixed languages”; e.g., Japanese, Tamil) organise rhythmic structure at a sub-syllabic level, with near-equal durations between moras--phonological units smaller than syllables (Ladefoged, 1975; Nespor et al., 2011). Since prosody encompasses both pitch variations and timing/duration variations, L1 rhythmic types may also have significant influence on prosody perception in non-native languages, with learners potentially perceiving prosody more easily in target languages that share their L1’s rhythmic typology. Despite this theoretical possibility, existing studies on the moderating role of L1 rhythmic type in music-language relationship are scarce, and no clear conclusions can be drawn at this point.

2.4.2 The Role of L2 Proficiency

As highlighted by Jansen et al. (2023) in their meta-analysis, proficiency of the target non-native language often functions as a continuous moderator in the music-prosody relationship. Proficiency spans a spectrum from complete unfamiliarity to near-native proficiency, with potentially distinct mechanisms potentially influencing prosody perception at different stages. Existing evidence suggests that, at the novice stage where the speaker is mostly unfamiliar with the target language, non-native perception is dominated by L1 transfer, and musical ability tends to be the strongest predictor of how well they can distinguish basic non-native sounds (Barbaroux et al., 2021); at the intermediate stage where the learner is more familiar with the language, higher proficiency in the target language improves prosody perception and mitigates, though not eliminates, L1 interference (Tremblay et al., 2016), though no existing study is known to have investigated the role of music abilities at this stage; at the advanced stage where learners reach near-native proficiency, they tend to demonstrate prosodic processing patterns approaching those of L1 speakers, with musical benefits limited to subtle pragmatic contrasts (e.g., emotional prosody, as seen in Mualem & Lavidor, 2015).

Beyond its moderating role in the music-prosody relationship, proficiency may also generally impact overall non-native prosody perception. For instance, studies on non-native prosody production demonstrate that lower proficiency learners perform worse than their higher proficiency counterparts (Gut, 2003). This finding may be extended to prosody perception, as inferred from Asano's (2018) study. Although not explicitly designed to assess the role of proficiency, Asano compared Japanese L1 speakers, L2 learners (i.e., higher proficiency) and non-learners (i.e., low proficiency) in discriminating Japanese consonant length contrasts under increased task demands. The results revealed a clear pattern: the less proficient the learners were, the more impaired their performance became under heightened cognitive load. This suggests that while proficiency may not eliminate all processing costs, it may significantly mitigate the impact of certain task demands on non-native prosody perception. However, a limitation of such studies is that their categorical groupings of proficiency make it challenging to delineate the influence of proficiency as spanning a continuous spectrum.

Despite these patterns and the general importance of proficiency, Jansen et al.'s (2023) review identified only 17% of studies on music-prosody transfer documenting proficiency measures, which leaves a critical gap in the literature and calls for future research to further investigate this mediating factor. Accordingly, this dissertation explicitly incorporates proficiency assessments as a continuous variable to address this gap, aiming to provide insights into how English proficiency influences the music-prosody link.

2.5 Research Gaps and Rationale for the Study

Though there is a general belief that music aptitude and non-native prosody perception are positively correlated, as suggested by a recent meta-analysis (Jansen et al., 2023), the substantial heterogeneity across studies included in it undermines this conclusion and warrants further exploration. Specifically, among the examined studies, there are significant variations in the specific type of musical ability assessed (innate aptitude vs. musicianship/ training), the granularity of prosody perceived (lexical vs. sentence level, pitch- vs. time-related), and participant demographics. Such inconsistencies in effect sizes, methodological designs, and measurement approaches across existing research point to critical gaps in the field.

This, combined with the methodological limitations in existing studies as discussed previously, all point to the need for a more rigorous approach. Specifically, previous studies suffer from: (1) the conflation of music aptitude and musical experiences, as the prevailing musician/non-musician dichotomous categorisation fails to disentangle innate music aptitude with acquired music experiences and overlooks the heterogeneity within musician populations. Moreover, while music training interventions offer insights into causality, their methodological diversity and frequent neglect of baseline individual differences in innate aptitude prevent robust conclusions on training effects. (2) Previous music aptitude tests often suffer from psychometric limitations, including a failure to truly separate innate aptitude from learned abilities, or exhibiting ceiling effects for neurotypical populations.

All these limitations underscore the necessity for studies employing psychometrically valid and reliable tests that measure the innate music aptitude as a continuum to rigorously investigate music aptitude’s impact on prosody perception. To address this research gap, this study adopts the psychometrically validated Mini-PROMS as the test battery. It not only measures music aptitude as a continuous numerical factor, but also allows for the analysis of how its distinct dimensions (melody, tuning, accent, tempo) individually predict specific prosodic subdomains (pitch and time).

Secondly, as suggested by theoretical models like the PAM and empirical evidence, L1 background can significantly affect L2/FL prosody perception and music perceptual abilities, which may potentially moderate the music-non-native prosody relationship. Specifically, different L1s that differ in the phonological distance from the target non-native language, or having different rhythmic types as the target language, may affect this relationship differently, a hypothesis that has not been tested by previous research. Therefore, this study regards language background primarily as a numerical factor by incorporating phonological distance metrics (L1-English distance scores spanning 0-1), which allows modeling of cross-linguistic transfer effects along a typological continuum, while supplementing this with investigation of rhythmic categorical differences between different L1s.

Thirdly, the role of target language familiarity in potentially mediating the relationship between musical abilities and non-native prosody perception remains inadequately explored.

Therefore, this study includes the explicit assessment of English proficiency as a continuous variable, thus enabling a detailed examination of how the music-prosody relationship varies across the competence spectrum, from limited to advanced learners.

Based on these identified research gaps and rationale, the present study seeks to answer the following research questions (RQs):

1. To what extent are music aptitude, L1, and language proficiency associated with prosody perception in a non-native language?

2. Does music aptitude predict prosody perception when controlling for L1 and language proficiency?

3. How do L1 and language proficiency moderate the relationship between musical aptitude and prosody perception?

4. Which aspects of prosody perception are most influenced by music aptitude, and how does this vary by L1 and language proficiency?

III. Methodology

This study employed a cross-sectional correlational design to examine the relationship between musical aptitude, L1 background, and non-native English prosody perception. A diverse sample of adult ESL/EFL learners with varying levels of English proficiency and different native language backgrounds was recruited. Participants were assessed on three primary measures: English proficiency using LexTALE, musical aptitude using the Mini Profile of Music Perception Skills (Mini-PROMS), and prosody perception using the Multi-Dimensional Battery of Prosody Perception (MBOPP). In addition, language distance between participants' L1 and the target language (English) was quantified using URIEL's genetic and featural datasets (Littell et al., 2017).

3.1 Participants

A priori power analyses were conducted using G*Power 3.1 (Faul et al., 2007) to determine the necessary sample size for the study to achieve 80% statistical power at an alpha level of 0.05. Based on Plonsky and Oswald's (2014) L2-specific effect size thresholds, the required sample sizes were as follows: For the correlation analysis (RQ1), a small-to-medium effect size (r=0.3) required a sample of N=84 participants. For the multiple regression model (RQ2), also using an effect size of r=0.3, the required sample size was N=77. For the linear mixed-effects models used in RQ3 and RQ4, which included five predictors and focused on detecting interaction terms, a small effect size (f^2=0.06) required a larger sample of approximately N=133 participants (see Appendix D for the detailed power analysis).

The power analyses collectively indicated a target sample size of 133 participants to ensure adequate power across all research questions. Due to practical constraints, a total of 100 participants were recruited for the study. This sample size was sufficient for addressing RQ1 and RQ2, but it was slightly underpowered for fully addressing the moderation effects in RQ3 and RQ4.

The participants were a diverse group of adult ESL/EFL learners with a range of L1 and English proficiency levels. The total sample of 100 participants included 49 females, 47 males, and 4 non-binary individuals. Participant ages ranged from 19 to 70 years, with a mean age of

31.64 years. A more detailed description of participant demography is presented in the following sections.

3.1.1 Linguistic Background

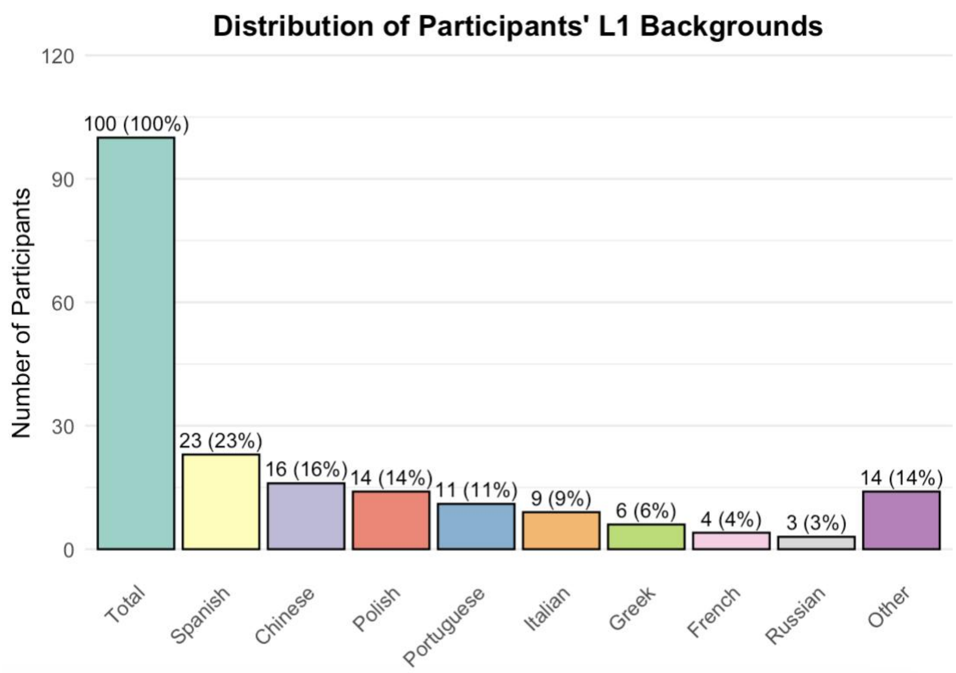
The participants represented a diverse range of linguistic backgrounds. All 100 participants reported speaking at least one other language in addition to their native language. Their majority of L1s included Spanish (n=23), Chinese (n=16), Polish (n=14), Portuguese (n=11), Italian (n=9), and Greek (n=6). The remaining 21 participants had 15 other L1 backgrounds, each with an individual count of 1 to 4. A detailed distribution of L1 backgrounds is presented in Figure 3.1.

Participants began learning English at an average age of 8.41 years (SD=5.27), with the age of acquisition ranging from 0 to 43 years old. The majority of participants reported a high level of proficiency in English, with 51 individuals self-rating their English as advanced/professional, 38 as intermediate or fluent, and 9 as native-like. Two participants rated their English proficiency as beginner. English usage in daily life was high, with 26 participants reporting they use English "Very often" (70-90%) and 17 reporting they use it "Almost always" (>90%).

Exposure to English in an English-speaking country was limited for most participants, with an average of 0.33 years (SD=0.47) spent in such environments.

Figure 3.1

Distribution of Participants' L1 Backgrounds



Note: L1s spoken by less than 3 participants were grouped into "Other", which includes: Afrikaans (n=1), Arabic (n=1), Bulgarian (n=1), Czech (n=1), Dutch (n=1), German (n=2), Hungarian (n=2), Korean (n=1), Turkish (n=1), Urdu (n=1), Urdu-Panjabi bilingual(n=1), Xhosa (n=1).

3.1.2 Educational and Musical Background

Regarding education, the highest level of achievement for the majority was a "Bachelor's degree or equivalent (n=43)", followed by a "Master's degree or equivalent" (n=30), and "Upper secondary education" (n=17).

The sample was primarily composed of non-musicians, with 58 participants identifying as "Music loving Non-musician" and 34 as "Non-musician". The remaining 8 participants were either "Amateur musician" (n=7) or "Semi-Professional musician" (n=1).

Music listening was a frequent activity, with 56 participants reporting they listen to music everyday. Out of the total participants, 22 reported playing a musical instrument or singing, with an average of 4.09 years of training (SD=6.67). Musical qualifications were uncommon, with only a few individuals reporting specific certifications or awards.

Table 3.1

Participant Demographics by Music Background

Music Background Category	N	Gender			Age			Education Level				Language Background
		Male	Female	Non-binary	19-30	31-50	>50	Upper Secondary	Bachelor	Master	Other	
Non-musician	34	17	16	1	10	17	7	6	15	12	1	Spanish (7), Chinese (6), Polish (4), Portuguese (3), Italian (3), Other (11)
Music loving Non-musician	58	28	28	2	25	28	5	9	24	16	9	Spanish (13), Chinese (7), Polish (6), Portuguese (8), Italian (5), Greek (3), Other (16)
Amateur musician	7	2	5	0	3	4	0	2	3	2	0	Spanish (2), Polish (2), Portuguese (1), Other (2)
Semi-Professional musician	1	0	0	1	0	1	0	0	1	0	0	Other (1)
Total	100	47	49	4	38	50	12	17	43	30	10	Spanish (23), Chinese (16), Polish (14), Portuguese (11), Italian (9), Greek (6), Other (21)

3.2 Materials

3.2.1 Music Aptitude

The Profile of Music Perception Skills (PROMS) (Law & Zentner, 2012) was used as the measurement for music aptitude in the current study. It has different versions with various test durations (10 to 60 minutes)—since the full version of PROMS takes an hour to finish and is thus lengthy, subsequent researchers have developed shortened versions of it, including the PROMS-S which takes 35 minutes to complete, the Mini-PROMS which takes 15 minutes, the Micro-PROMS which takes 10 minutes, and the Modular-PROMS, which allows flexible selection among its 9 subtests (Zentner & Strauss, 2017). All versions have been tested to have good validity and reliability.

The current study uses the Mini-PROMS version for several reasons. First, it only takes 15 minutes to complete and thus is time efficient. Second, it consists of four subtests: Melody, Tuning, Accent (i.e., rhythmic patterns), and Tempo (i.e., speed). These subtests were selected from the full version because of their strong association with the full-length PROMS and because of their robustness against acoustic variations in testing environments (Zentner & Strauss, 2017). In addition, these subtests also provide separate reports of participants' tonal and rhythmic abilities, which enables the analysis of how different aspects of music aptitude predict different aspects of prosody perception (i.e., pitch and timing).

The Mini-PROMS is publicly accessible from its official website (https://webapp.uibk.ac.at/psychologie/psyuibk/index.php/163871?lang=en). A link to this website was embedded in Gorilla, the online experiment-building platform.

3.2.2 English Prosody Perception

This study uses the Multi-Dimensional Battery of Prosody Perception (MBOPP) (Jasmin et al., 2020) to access participants' perception of prosody in English sentences. MBOPP is chosen for several reasons: Firstly, it is targeted at typically developing adults and, therefore, does not suffer from ceiling effects that many existing assessments of prosodic perception have. Second, it can examine participants' perception of individual acoustic dimensions (pitch and time) separately and distinguish the unique contributions of individual acoustic features of pitch and temporal cues. Third, it consists of two subtests, one on linguistic focus - which measures the ability to hear emphasis or sentential stress, and the other on phrase boundaries - which measures the ability to hear where in a compound sentence one phrase ends and another begins. This allows for a more comprehensive assessment of prosody perception, capturing different aspects of prosodic processing. Finally, the test difficulty can be precisely controlled as the stimuli were created using a continuous voice morph space and a set of stimuli with varying difficulty were provided by the test developers and can be downloaded from GitHub.

In the current study, MBOPP stimuli were downloaded from GitHub and the test was manually built through the online experiment-building platform Gorilla. In part 1 of the test (the linguistic focus subtest), there are 44 test trails in total, half of them were set in the

Pitch-Only condition, where the duration (time) morphing rates were held at 50%, while two contrasting pitch versions were controlled at 0% (towards early focus) and 100% (towards late focus). This is different from the original test where the contrasting pitch were held at 25% and 75%. This is because that the original test was targeted at English native speakers, while the current study targets at L2 and FL speakers of English. Thus, this study adopts the easiest version of stimuli to avoid floor effects. Similarly, in the Time-Only condition, pitch was held at 50% while duration was either 0% (early focus) or 100% (late focus). The stimuli consist of 25 compound sentences in total (1 for example, 2 for practices, and 22 for the test).

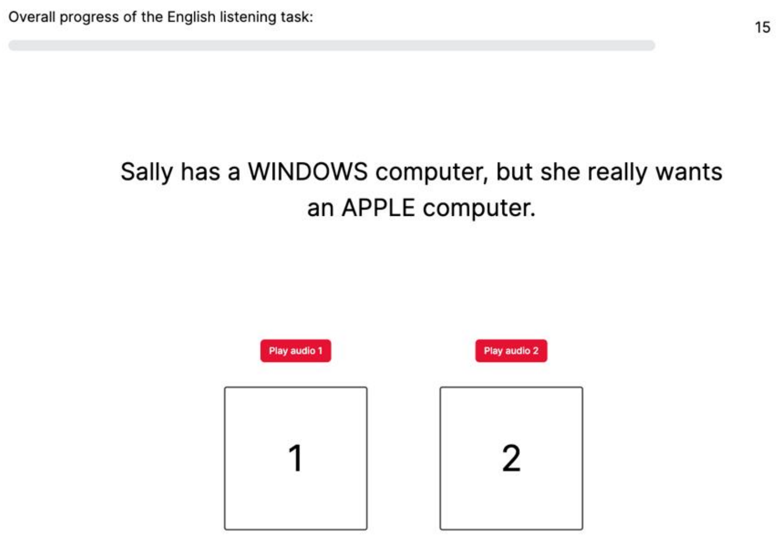
In part 2 of the test (the phrase boundary perception subtest), there are 40 test trails in total. Half of them were set in the Pitch-Only condition, where the duration (time) morphing rates were held at 50%, while two contrasting pitch versions were controlled at 0% (towards early closure) and 100% (towards late closure). Similar to part 1, the stimuli in this part were also adapted from the original test to lower its difficulty level so that it would be more suitable for non-native speakers of English. The stimuli in this part consist of 23 compound sentences in total (1 for example, 2 for practices, and 20 for the test). The full stimuli sentences can be seen in Appendix F.

In part 1 of the test, participants would first see a sentence with one part in all upper-case letters indicating emphasis. They will then hear two recordings of the first half of the sentence, each with different emphasis. Since the second half of the sentence may change according to the different emphasis, the second half would not be read aloud in the recordings. Only one recording corresponds to the written sentence, and the participants need to choose which one is the correct answer by pressing “1” or “2” on the screen. They have 20 seconds to answer each test trial. Similarly, in part 2, they would also see a written sentence and hear two recordings of the first half of the sentence, each with different pauses. The two recordings only differ in pauses (i.e., early closure or late closure as indicated by the comma in the sentence), and participants need to choose which one corresponds to the written sentence by pressing “1” or “2” on the screen. Before each test section, they will see an example and try two practice trails to ensure full understanding of the task.

In both subtests, the order of test trails was randomised to minimise potential order effects.

Figure 3.2

Example of a MBOPP Part I test trail



Note. Participants would hear the two audio recording of the first half of the sentence "Sally has a windows computer" and would have to pick the appropriate response.

3.2.3 English Proficiency Level

The English version of LexTALE (Lemhöfer & Broersma, 2012) was chosen to assess participants' English proficiency in this study due to its reliability, efficiency, and suitability for measuring vocabulary knowledge, which is a strong indicator of overall language proficiency (Lemhöfer & Broersma, 2012). LexTALE is a quick and simple lexical decision task, is freely available, easy to administer online, and requires minimal time to complete, reducing participant fatigue. Additionally, Lemhöfer & Broersma (2012) also reported that LexTALE scores correlate well with standardised test results of language proficiency such as TOEFL and IELTS, demonstrating its validity as an effective assessment tool. Its objective scoring system also minimises potential biases associated with self-reported proficiency levels.

When developing the test materials using Gorilla, LexTALE was available in Gorilla's Open Materials and was cloned to the current project. To ensure clarity for participants, an additional instruction—"Decide whether it is a real word in English or not."—was added to each test trial. Furthermore, the original progress bar was replaced with a more specific progress cue—"Question n/60"—as the default progress bar in Gorilla is designed to track the entire experimental session rather than individual tasks.

3.2.4 Background Questionnaire

A background questionnaire was also developed in Gorilla. It collected participants' basic information (gender, age, nationality), educational background, language background and language exposure, and music background. These information provide essential context for the potential moderating variables in this study. The complete questionnaire can be seen in Appendix C.

3.3 Procedure

3.3.1 Piloting

Prior to data collection, pilot testing was conducted with five participants from the Department of Education at the University of Oxford. Based on their feedback, several adjustments were made to the experiment. The most significant change was reducing the difficulty level of the prosody perception test by modifying the stimuli, i.e., by using more distinct pitch and duration contrasts (0% vs. 100%) instead of the original 25% vs. 75%, making the stimuli easier for L2 and FL English learners. Additional refinements included renaming tasks to be more participant-friendly, clarifying task instructions, adjusting time limits for each section, revising wording and response options in the background questionnaire, incorporating encouraging messages at the end of each section to promote participant engagement, and resolving technical issues such as repositioning buttons for better usability and adding a progress bar specifically for the prosody perception test.

3.3.2 Administration

After developing the whole experiment, the experiment link was generated using the Gorilla platform. Participants were primarily recruited via Prolific.com, a participant recruitment website that enables pre-screening based on specified criteria. The pre-screener conditions applied in this study included "First language not English," "Bilingual," "No language-related disorders," and "Age 18–99."

Following this initial recruitment, a demographic review revealed a majority of participants with European linguistic backgrounds. To enhance the sample's linguistic diversity, particularly by including participants with Asian language backgrounds, supplementary recruitment was conducted through personal contacts via the Chinese social media platform WeChat. Potential participants were informed about the study through a brief overview posted on WeChat Moments. Interested individuals then messaged the researcher and were subsequently provided with detailed instructions and the Gorilla experiment link. Participants were granted two to three days to complete the tasks at their convenience. The experiment content delivered via the Gorilla link was identical across both recruitment channels. Participant progress was monitored through the Gorilla backend, which displayed anonymised participant IDs and their current status (e.g., live, completed, rejected due to time limit, or manually rejected). Recruitment was manually concluded once the target sample size of completed participants was achieved.

3.3.3 Experiment Procedure

Participants were first presented with the CUREC approval information sheet and the consent form (see Appendix B). To participate in the study, participants had to declare their willingness by ticking the boxes "I certify that I'm 18 years of age or over." and "Yes, I agree to take part." to continue. Upon consenting, participants completed the experimental tasks in a fixed sequence: the music aptitude test (Mini-PROMS), the English prosody perception test (MBOPP), the English proficiency test (LexTALE), and the background questionnaire.

The Mini-PROMS test was completed on its publicly accessible official website (https://webapp.uibk.ac.at/psychologie/psyuibk/index.php/163871?lang=en), to which participants were redirected from the Gorilla platform. Upon finishing the test, participants received a detailed report of their music aptitude scores (total and four subtests: melody, tuning, accent, and tempo). They were then instructed to return to the Gorilla platform to manually enter these results. Progression to subsequent experimental sections was contingent upon the completion of this step. Following this, participants completed the MBOPP, LexTALE, and the background questionnaire. The entire experiment typically took between 50 and 90 minutes to complete. All data collection procedures adhered to CUREC guidelines, ensuring data security and confidentiality.

3.3.4 Ethical Considerations

Data collection and management were conducted in accordance with the University of Oxford's Data Protection and Research guidelines (Central University Research Ethics Committee, 2018). Ethical approval for the online experiment was obtained prior to participant recruitment (see Appendix A). In line with fundamental ethical principles such as anonymisation, informed consent, and transparent information provision (Tao et al., 2017), participants were supplied with consent and information forms before participation. All data gathered were anonymised through the secure encryption systems embedded within both Gorilla and Prolific.com platforms, ensuring that no personally identifiable information was accessible. Following the CUREC Best Practice Guidance 06 for Internet-mediated Research (Central University Research Ethics Committee, 2021), all digital data—including consent documentation and processed datasets used for coding and statistical analyses—were securely stored in a password-protected file on an external hard drive.

3.4 Data Analysis

All statistical analyses were conducted using R statistical software. Before proceeding with inferential tests, preliminary normality checks were performed (including Shapiro-Wilk tests, skewness, kurtosis, and visual inspections). These checks revealed that several key variables (Prosody Total, Prosody Pitch Total, Prosody Time Total, and L1 Distance) exhibited non-normal distributions. Consequently, non-parametric statistical methods and bootstrapping procedures were extensively employed to ensure the robustness of the findings. All continuous predictor variables were z-standardised prior to their inclusion in regression models.

3.4.1 Data Analysis for RQ1 (Associations)

To examine the associations between music aptitude, L1 background, English proficiency, and non-native prosody perception (RQ1), Spearman's rank-order correlation coefficients (ρ) were calculated. A Bonferroni correction was applied to the significance level to adjust for multiple comparisons.

3.4.2 Data Analysis for RQ2 (Prediction)

For RQ2, which investigated whether music aptitude predicts prosody perception while controlling for L1 and language proficiency, a multiple linear regression model was used. Given the non-normal distribution of residuals, bootstrapping with 1000 replicates was employed to obtain robust standard error estimates and 95% percentile confidence intervals for the regression coefficients.

3.4.3 Data Analysis for RQ3 (Moderation)

To explore the moderating roles of L1 background and English proficiency on the music aptitude-prosody perception relationship (RQ3), a moderated multiple regression model was conducted. This model included main effects and interaction terms for PROMS Total × L1 Distance and PROMS Total × LexTALE. Bootstrapping with 1000 replicates was again applied to generate robust 95% percentile confidence intervals for the interaction terms.

3.4.4 Data Analysis for RQ4 (Specific Aspects and Moderation)

To address RQ4, a series of multiple linear regression models were run. This involved testing how specific music aptitude subtests (Melody, Tuning, Accent, Tempo) predicted pitch-related and timing-related prosody skills, along with their interactions with L1 language distance and English proficiency. In total, 8 separate models were built (4 PROMS subtests for each of the 2 prosody dimensions). Bootstrapping with 1000 replicates was used in all these models to ensure robust confidence intervals for all coefficients.

3.4.5 Follow-up Analyses for L1 as a Categorical Variable

Finally, to further explore the influence of L1 background, follow-up regression analyses were conducted for all research questions where L1 was re-coded and treated as a categorical variable (stress-timed, syllable-timed, and mixed/mora-timed languages). Similar to the primary analyses, bootstrapping with 1000 replicates was consistently used to obtain robust confidence intervals for coefficients in these models.

IV. Results

This chapter first presents the descriptive analysis of the data, followed by the results for each research question (RQ). The descriptive analysis includes the distribution characteristics of the four key variables and normality assessments. These analyses provide a foundation for subsequent inferential statistical tests and ensure the appropriateness of the selected statistical methods. As initial normality checks indicated deviations from normal distribution for several key variables, non-parametric analyses were subsequently employed to address each research question.

The study's four key variables are music aptitude scores, prosody perception scores, English proficiency scores, and L1 language distance. The music aptitude scores are further differentiated into a total score (coded as *PROMS Total*) and four subtest scores (coded as *PROMS Melody*, *PROMS Tuning*, *PROMS Accent*, and *PROMS Tempo*). Similarly, prosody perception scores comprise a total score (coded as *Prosody Total*), total scores for the two conditions (coded as *Prosody Pitch Total* and *Prosody Time Total*), total scores for the two subtests (coded as *Prosody Focus Total* and *Prosody Phrase Total*), and individual scores for each condition within the two subtests (coded as *Prosody Focus Pitch, Prosody Focus Time, Prosody Phrase Pitch*, and *Prosody Phrase Time*). The English proficiency score was coded as *LexTALE*, and language distance was coded as *L1 Distance*.

4.1 Descriptive Analysis

This section provides a comprehensive overview of the central tendency, dispersion, and range of scores for all key variables in the study, including total scores and their respective sub-components. These descriptive statistics clearly demonstrate the characteristics of the participant sample and their performance on the various test measures.

Table 4.1, Table 4.2 and Table 4.3 presents the descriptive statistics for the key variables: Total scores, including *PROMS Total* (music aptitude), *Prosody Total* (overall prosody perception), *Prosody Pitch Total* (pitch-related prosody perception), *Prosody Time Total* (time-related prosody perception), *LexTALE* (English proficiency), and *L1 Distance* (L1 language distance), and subtest scores of both music aptitude and prosody perception, as Table 4.3 presents.

Table 4.1

Descriptive Statistics of Total Scores

Variable	N	M	SD	SE	Median	Min	Max	Range	Skew	Kurtosis	z_skew	z_kurtosis
PROMS Total	100	20.795	4.93	0.49	21	9.5	32	22.5	-0.18	-0.44	-0.73	-0.90
Prosody Total	100	70.33	8.92	0.89	71	43	84	41	-0.94	0.81	-3.83	1.66
Prosody Pitch Total	100	34.58	5.09	0.52	35	21	42	21	-0.50	-0.42	-2.02	-0.85
Prosody Time Total	100	35.75	4.59	0.46	37	20	42	22	-1.25	1.80	-5.12	3.68
LexTALE	100	48.8	6.27	0.63	49	31	59	28	-0.36	-0.39	-1.48	-0.80
L1 Distance	100	0.42	0.13	0.01	0.34	0.195	0.5687	0.37	-0.049	-1.62	-0.2	-3.31

Note. The maximum scores are: 36 for PROMS Total, 84 for Prosody Total (42 for Prosody Pitch Total and 42 for Prosody Time total), and 60 for LexTALE.

Table 4.2

Descriptive Statistics of Music Aptitude Test (PROMS) Scores

Variable	N	M	SD	SE	Median	Min	Max	Range	Skew	Kurtosis	z_skew	z_kurtosis
PROMS Total	100	20.80	4.93	0.49	21	9.5	32	22.5	0.18	0.44	0.73	0.90
PROMS Melody	100	5.28	1.79	0.18	5	1.5	9	7.5	0.07	0.56	0.28	1.15
PROMS Tuning	100	4.32	1.54	0.15	4	1	8	7	0.10	0.47	0.43	0.96
PROMS Accent	100	5.46	1.84	0.18	5.5	1.5	9	7.5	0.01	0.84	0.02	1.72
PROMS Tempo	100	5.74	1.51	0.15	6	1.5	10	8.5	0.15	0.05	0.61	0.09

Figure 4.1

Test Developers' Descriptive Statistics of Mini-PROMS (Zentner & Strauss, 2017)

	Raw scores		d'		Reliability	
Subtest	*M*	SD	*M*	SD	ω	$r_{tt}{}^{a}$
Melody	6.43	1.81	0.64	0.21	0.71	0.68^{***}
Tuning	5.44	1.55	0.91	0.33	0.76	0.63^{***}
Tempo	6.15	1.35	1.10	0.26	0.53	0.51^{**}
Accent	6.50	1.67	0.73	0.24	0.63	0.69^{***}
Mini-PROMS	24.56	4.78	0.68	0.10	0.82	0.83^{***}

${}^{a**}P < 0.01$; ${}^{***}P < 0.001$.
$n = 150$; $n^a = 31$.

Table 4.3

Descriptive Statistics of Prosody Perception Test Scores

Variable	N	M	S.D.	S.E.	Median	Min	Max	Range	Skew	Kurtosis	z_skew	z_kurtosis
Prosody Total	100	70.33	8.92	0.89	71	43	84	41	0.94	0.81	3.83	1.66
Prosody Pitch Total	100	34.58	5.09	0.51	35	21	42	21	0.50	0.42	2.02	0.85
Prosody Time Total	100	35.75	4.59	0.46	37	20	42	22	1.25	1.80	5.12	3.68
PROMS Focus Pitch	100	18.71	3.49	0.35	20	8	22	14	0.99	0.02	4.03	0.04
PROMS Focus Time	100	18.67	2.89	0.29	19	9	22	13	1.21	1.56	4.94	3.18
PROMS Focus Total	100	37.38	5.62	0.56	39	17	44	27	1.26	1.59	5.14	3.24
PROMS Phrase Pitch	100	15.87	3.02	0.30	16	9	20	11	0.48	0.69	1.98	1.41
PROMS Phrase Time	100	17.08	2.47	0.25	18	8	20	12	1.22	1.26	4.99	2.58
PROMS Phrase Total	100	32.95	5.01	0.50	34	17	40	23	0.86	0.29	3.51	0.60

From the tables, the mean ***PROMS Total*** score of 20.795 out of 36 indicated a moderate level of general music aptitude within the sample, with noticeable individual differences. A comparison with the test developer-reported data (Figure 4.1) revealed that while the standard deviations were largely comparable, the current study's participants achieved consistently lower mean scores across the total PROMS and its subtests, suggesting that the current sample may represent a different demographic background compared to the developer's sample. For prosody perception, participants generally performed well, with a mean *Prosody Total* score of 70.33 out of 84, suggesting overall proficiency in the target language, English, which is also consistent with the *LexTALE* results which indicated an intermediate to upper-intermediate level among participants, with a mean of 48.8 out of 60. Within prosody, performance was slightly better on time-related sub-components compared to pitch-related ones. Finally, L1 language background results confirmed a diverse range of L1 backgrounds and linguistic distances from English within the study's participant cohort.

Prior to statistical analysis, normality checks were conducted to validate the assumptions underlying parametric statistical tests (see Appendix H). Results showed that *PROMS Total* and *LexTALE* exhibit normal distributions, while *Prosody Total*, *Prosody Pitch Total*, *Prosody Time Total*, and *L1 Distance* demonstrate significant deviations from normality. These findings necessitate the use of non-parametric statistical methods for addressing the research questions with variables that are not normally distributed.

Following normality assessment, outlier detection was performed for continuous variables, and four outliers within the Prosody Total scores were identified (see Appendix G). However, these data points were retained in the analysis, as these scores represent genuine individual differences in prosody perception rather than data entry errors or measurement flaws (Field et al., 2012). Their retention thus ensures the subsequent analyses encompass the full range of individual variation present in the sample. Given the study's focus on individual differences and the use of non-parametric statistical methods and bootstrapping procedures (which are robust to deviations from normality and the presence of outliers), the inclusion of all data points was considered appropriate.

4.2 Data Analysis for RQ1: To what extent are music aptitude, L1, and language proficiency associated with prosody perception in a non-native language?

Given that the variables *Prosody Total* and *L1 Distance* exhibited non-normal distributions, Spearman's correlation coefficient (ρ), a non-parametric statistic, was employed to address RQ1. To account for multiple comparisons, a Bonferroni correction was applied, setting the new significance threshold at $p = 0.0167$ (0.05/3). The results are summarised in Table 4.4.

Table 4.4

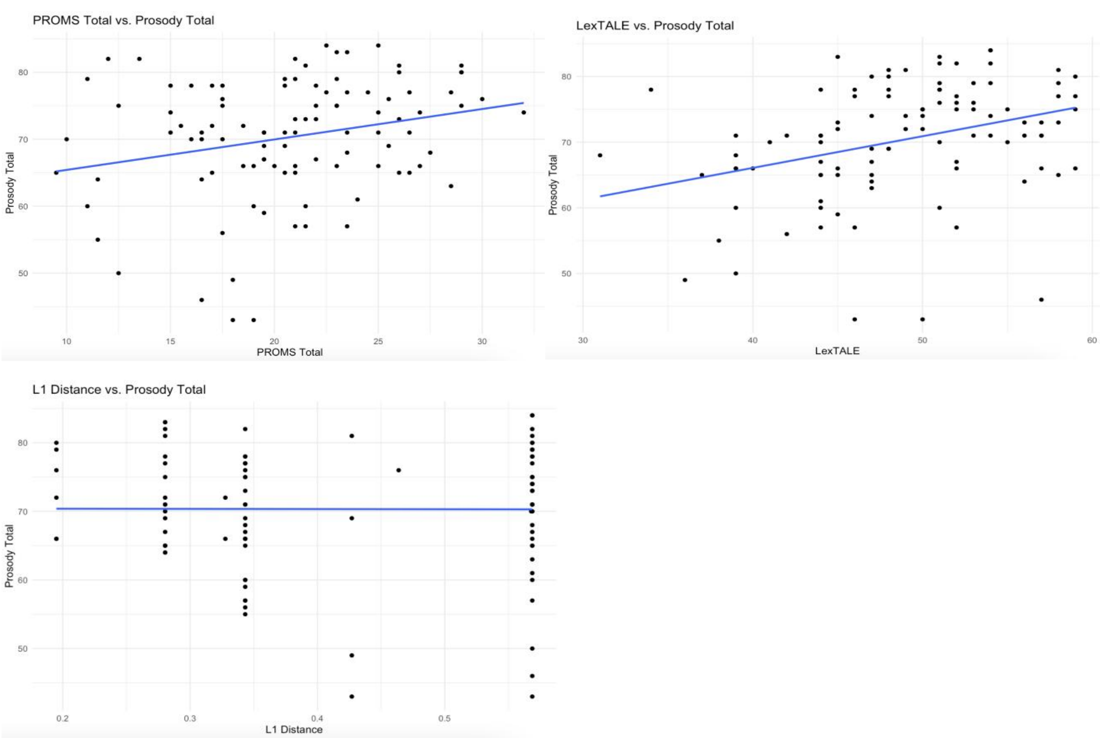
Spearman Correlation Results for RQ1 (Total Scores)

Variable 1	**Variable 2**	**Spearman's ρ**	**p_value**
PROMS Total	*Prosody Total*	0.244	0.0146*
L1 Distance	*Prosody Total*	0.032	0.7488
LexTALE	*Prosody Total*	0.389	0.0001*

Note. Significance at p<0.0167, marked with an asterisk (*).

Figure 4.2

Scatter Plots of Spearman Correlation Results for RQ1 (Total Scores)



As shown, a significant positive correlation was found between *PROMS Total* and *Prosody Total*, with a small to moderate effect (ρ=0.244) according to Cohen's (1988) guidelines (0.10: small effect; 0.30: moderate effect; 0.50: large effect). Similarly, a significant positive correlation with a moderate effect was observed between *LexTALE* and *Prosody Total*, indicating a stronger association than that of music aptitude. In contrast, no statistically significant relationship was found between L1 distance and prosody perception.

4.3 Data Analysis for RQ2: Does music aptitude predict prosody perception when controlling for L1 and language proficiency?

To address RQ2, multiple linear regression was conducted. This model aimed to identify the unique predictive contributions of each independent variable to prosody perception. The dependent variable (DV) was *Prosody Total*, while the independent variables (IVs) were *PROMS Total*, *L1 Distance*, and *LexTALE*.

Before interpreting the regression coefficients, the assumption of normality of the multiple linear regression model was examined. The distribution of residuals were visually inspected through a histogram and a Q-Q plot (details can be seen in Appendix H), which indicated a deviation from normality. This necessitates using bootstrapping, a powerful, non-parametric resampling technique that does not require assumptions about data distribution like normality (Field et al., 2012), as a supplementary method for robust inference.

The results of the standardised multiple linear regression model are summarised in Table 4.8. The overall regression model was statistically significant ($F(3,96)=6.488$, $p=0.00048$), explaining 16.86% of the variance in non-native prosody perception accuracy ($R^2=0.1686$). This suggests that the included predictors collectively contribute meaningfully to understanding prosody perception.

Table 4.5

Standardised Multiple Linear Regression Results for RQ2

Variable	β (Std. Estimate)	Std. Error	Cohen's f^2	95% Percentile CI (Bootstrapped)
(Intercept)	NA	7.076	-	-
PROMS Total	0.2317	0.17	0.06	(0.1017, 0.7588)*
L1 Distance	-0.1006	6.551	0.01	(-20.215, 5.534)
LexTALE	0.3239	0.138	0.12	(0.2074, 0.7575)*
Model Statistics:	**R^2**	**Adjusted R^2**	**F-statistic**	**p-value**
	0.1686	0.1426	6.488 (df =3,96)	0.0004808*

Note. A 95% bootstrapped confidence interval that does not contain zero indicates a significant effect, marked with an asterisk (*).

Figure 4.3

Histogram of Standardised β for RQ2 Regression

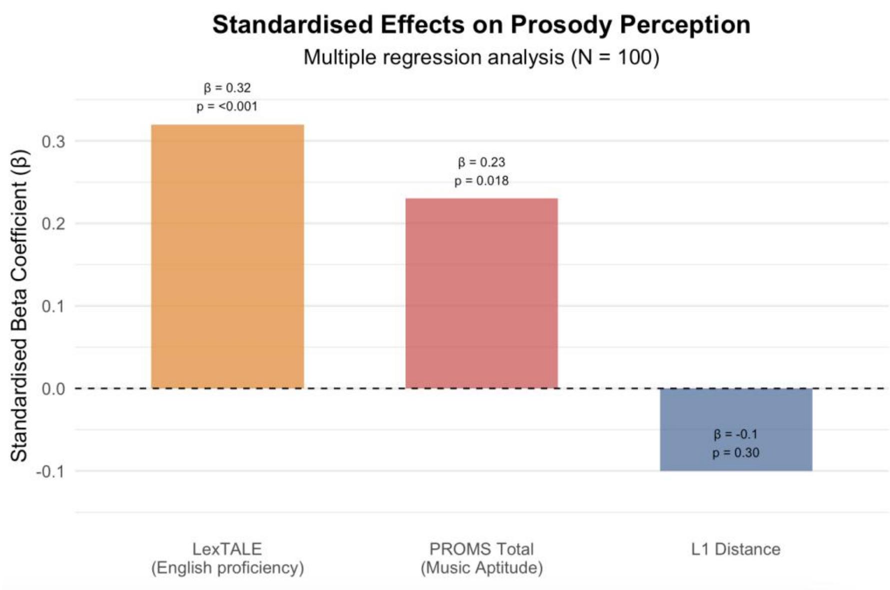
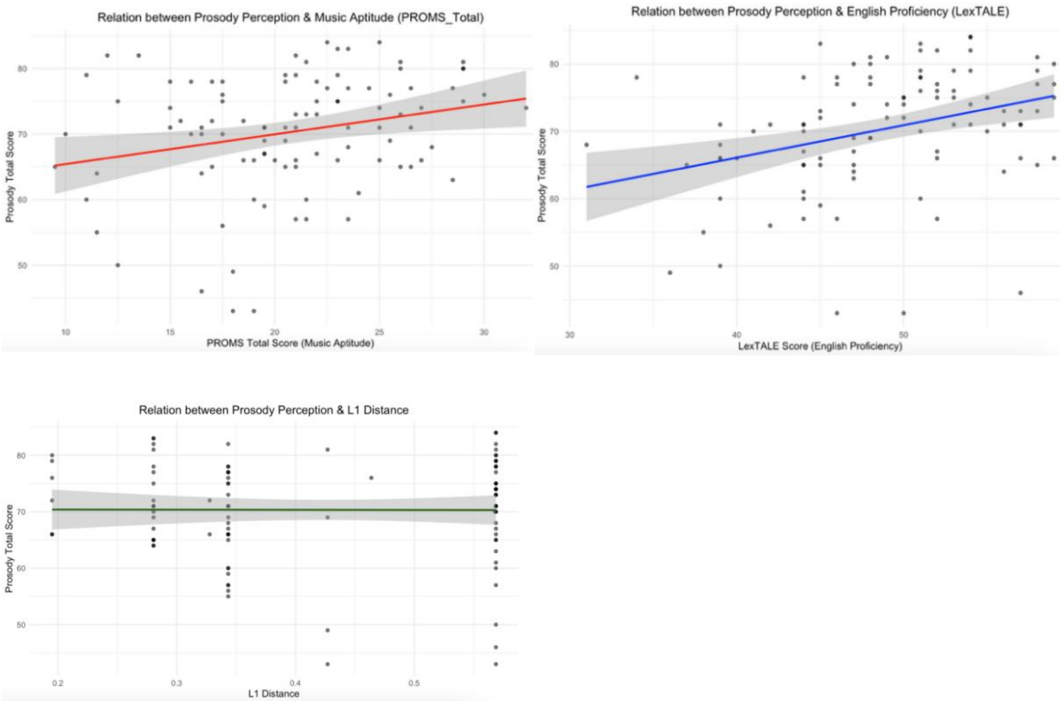


Figure 4.4

Scatter Plots of Regression Results for RQ2



Given the non-normal distribution of residuals, a bootstrapping procedure with 1000 replicates was performed to obtain robust standard errors and confidence intervals for the regression coefficients. Based on the bootstrapped regression results, individual predictor analysis revealed the following:

Overall music aptitude (*PROMS Total*) was found to be a significant positive predictor of prosody perception accuracy, as its 95% bootstrapped confidence interval (0.1017, 0.7588) does not contain zero. The Cohen's f^2 of 0.06 indicates a small-to-moderate effect size, suggesting that music aptitude provides a small contribution to explaining variance in prosody perception.

LexTALE was also a highly significant positive predictor of non-native prosody perception accuracy, as its 95% bootstrapped confidence interval (0.2074, 0.7575) does not contain zero. The Cohen's f^2 of 0.12 suggests a moderate effect size, indicating that higher English proficiency, as measured by vocabulary knowledge, is strongly associated with better prosody perception. This finding positions LexTALE as the most prominent predictor in the model, demonstrating a stronger predictive power than music aptitude.

L1 language distance, however, did not emerge as a statistically significant predictor of non-native prosody perception accuracy, as its 95% bootstrapped confidence interval (-20.215, 5.534) includes zero. The negative standardised beta value ($\beta = -0.101$) suggests a non-significant trend where greater L1 phonological distance is associated with poorer non-native prosody perception.

In summary, the multiple linear regression analysis for RQ2, robustly supported by bootstrapping, indicates that both English proficiency (*LexTALE*) and overall music aptitude (*PROMS Total*) are significant positive predictors of non-native prosody perception accuracy. English proficiency exhibited the strongest predictive power, followed by music aptitude. L1 language distance, however, was not found to be a significant predictor in this model. This finding is fully consistent with the non-significant correlation between L1 language distance and prosody perception found in the previous analysis. While the model significantly explained 16.86% of the variance in prosody perception, this relatively small R^2 suggests that a substantial portion of variance remains unexplained, pointing to the potential influence of additional predictors.

4.4 Data Analysis for RQ3: How do L1 and language proficiency moderate the relationship between musical aptitude and prosody perception?

To address RQ3, a multiple regression model was built. The DV was *Prosody Total, and predictors included PROMS Total*, *L1 Distance*, and *LexTALE*, along with two interaction terms: *PROMS Total*× *L1 Distance* and *PROMS Total*× *LexTALE*. All continuous predictor variables were z-standardised before creating the interaction terms to facilitate interpretation and reduce multicollinearity.

The overall regression model was statistically significant ($F(5,94)=3.977$, $p=0.002569$), explaining 17.46% of the variance in non-native prosody perception accuracy ($R^2=0.1746$).

Table 4.6

Standardised Moderated Regression Results for RQ3

Variable	β (Std. Estimate)	Std. Error	95% Percentile CI (Bootstrapped)
(Intercept)	70.4936	0.857	(68.63, 72.15)*
PROMS Total	2.1364	0.8688	(0.593, 3.892)*
L1 Distance	-0.9061	0.8638	(-2.5852, 0.7865)
LexTALE	2.8887	0.8488	(1.221, 4.567)*
PROMS Total: L1 Distance	-0.4378	0.8685	(-1.9883, 1.3775)
PROMS Total: *LexTALE*	-0.5386	0.9208	(-2.4195, 1.7113)
Model Statistics:			
R^2	0.1746		
Adjusted R^2	0.1307		
F-statistic	3.977 (df = 5, 94)		
p-value	0.002569*		

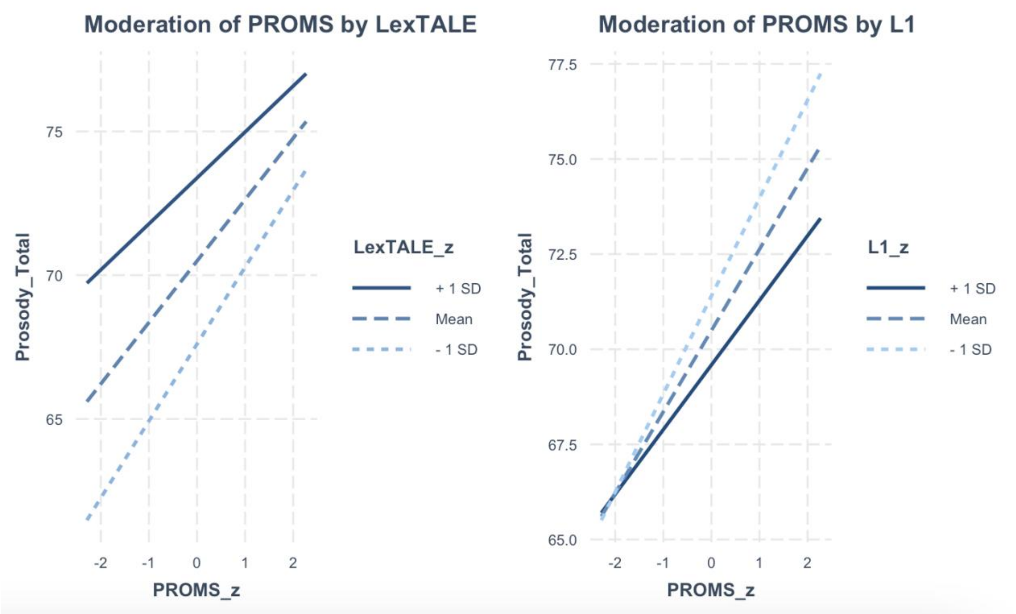
Note. A 95% bootstrapped confidence interval that does not contain zero indicates a significant effect, marked with an asterisk (*).

The main effects of the predictors in this model showed a similar pattern of significance to the previous model, which did not include interaction terms. Both music aptitude (β=2.1364, 95% CI: 0.593, 3.892) and English proficiency (β=2.8887, 95% CI: 1.221, 4.567) continued to show significant positive effects on prosody perception, as their confidence intervals did not include zero. In contrast, L1 distance (β=−0.9061, 95% CI: -2.5852, 0.7865) remained a non-significant predictor.

The analysis of the interaction terms revealed that the relationship between music aptitude and prosody perception was not significantly moderated by either L1 language distance or English proficiency. The bootstrapped 95% confidence interval for the *PROMS Total: L1 Distance* interaction term (-1.9883, 1.3775) includes zero, indicating that L1 distance presented a non-significant effect. Similarly, the confidence interval for the *PROMS Total: LexTALE* interaction term (-2.4195, 1.7113) also includes zero, confirming proficiency's lack of a significant moderating role. However, while the interaction between music aptitude and L1 distance was not statistically significant, visual inspection of the moderation plot (Figure 4.8, right panel) suggests a non-significant trend where the positive predictive power of music aptitude on prosody perception appears slightly stronger when a learner's L1 is phonologically closer to the target language (represented by lower *L1 Distance* values).

Figure 4.5

Visualisation of Interaction Results for RQ3



Note. PROMS_z, LexTALE_z, and L1_z represent the z-standardised scores for PROMS Total (overall music aptitude), LexTALE (English proficiency), and L1 Distance (L1 language distance), respectively, which corresponds to the variables in Table 4.9.

In summary, both music aptitude and English proficiency continued to show significant positive main effects on prosody perception. However, neither English proficiency nor L1 language distance was found to significantly moderate the relationship between music aptitude and prosody perception. The visual inspection, though, suggests a potentially stronger influence of music aptitude for learners with L1 phonologically closer to English, though this effect did not reach statistical significance.

4.5 Data Analysis for RQ4: Which aspects of prosody perception are most influenced by music aptitude, and how does this vary by L1 and language proficiency?

To address RQ4, a series of multiple linear regression models were conducted. This approach is similar to the one used for RQ3, but here the DV was changed to reflect specific aspects of prosody perception, while the IVs were changed to the specific music aptitude subtest scores. Separate models were run with two distinct DVs: *Prosody Pitch Total* and *Prosody Time Total*. The IVs included *PROMS Melody*, *PROMS Tuning*, *PROMS Accent*, and *PROMS Tempo*, while *L1 Distance* and *LexTALE* were included as main effects and as moderators via interaction terms. All continuous variables were z-standardised prior to analysis. In total, 8 separate regression models were run. Each model tested one of the four PROMS subtests as the primary predictor for either pitch-related or timing-related prosody skills. The bootstrapped 95% confidence intervals (CIs) for the standardised regression coefficients (β) are summarised in Table 4.10 and visually presented in a heatmap (Figure 4.9). Each row in Table 4.7 represents the output of a single regression analysis.

Table 4.7

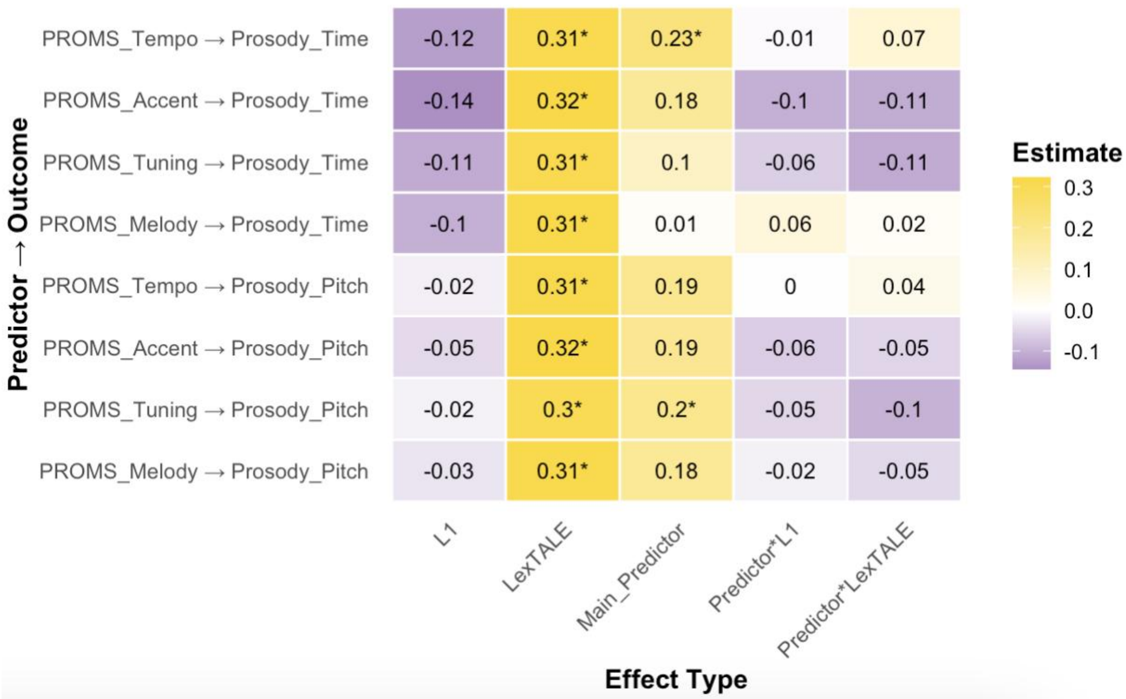
Standardised Regression Coefficients (β) for RQ4 Models Predicting Prosody Perception

Predictor → Outcome	L1 Estimate (95% CI)	LexTALE Estimate (95% CI)	Main Predictor Estimate (95% CI)	Predictor × L1 Estimate (95% CI)	Predictor × LexTALE Estimate (95% CI)
PROMS Melody → Prosody_Pitch	-0.03 (-0.21, 0.17)	0.31 (0.12, 0.51)*	0.18 (-0.03, 0.39)	-0.02 (-0.27, 0.21)	-0.05 (-0.31, 0.19)
PROMS Tuning → Prosody_Pitch	-0.02 (-0.19, 0.19)	0.30 (0.12, 0.49)*	0.20 (0.00, 0.39)*	-0.05 (-0.25, 0.14)	-0.10 (-0.34, 0.16)
PROMS Accent → Prosody_Pitch	-0.05 (-0.24, 0.15)	0.32 (0.14, 0.52)*	0.19 (-0.01, 0.42)	-0.06 (-0.28, 0.16)	-0.05 (-0.22, 0.13)
PROMS Tempo → Prosody_Pitch	-0.02 (-0.22, 0.16)	0.31 (0.14, 0.51)*	0.19 (-0.02, 0.41)	0.00 (-0.22, 0.21)	0.04 (-0.26, 0.25)
PROMS Melody → Prosody_Time	-0.10 (-0.31, 0.09)	0.31 (0.12, 0.53)*	0.01 (-0.19, 0.22)	0.06 (-0.19, 0.26)	0.02 (-0.18, 0.30)
PROMS Tuning → Prosody_Time	-0.11 (-0.31, 0.07)	0.31 (0.11, 0.51)*	0.10 (-0.11, 0.30)	-0.06 (-0.25, 0.16)	-0.11 (-0.38, 0.21)
PROMS Accent → Prosody_Time	-0.14 (-0.33, 0.03)	0.32 (0.12, 0.52)*	0.18 (-0.01, 0.37)	-0.10 (-0.31, 0.10)	-0.11 (-0.27, 0.11)
PROMS Tempo → Prosody_Time	-0.12 (-0.32, 0.05)	0.31 (0.12, 0.54)*	0.23 (0.06, 0.41)*	-0.01 (-0.19, 0.17)	0.07 (-0.16, 0.28)

Note. A 95% bootstrapped confidence interval that does not contain zero indicates a significant effect, marked with an asterisk (*). Estimates are rounded to two decimal places for readability in the table.

Figure 4.6

Standardised Regression Coefficients for PROMS Subtests Predicting Prosody



Based on the regression analyses, several key findings emerged regarding the influence of specific music aptitude dimensions on prosody perception and the moderating roles of L1 and English proficiency.

4.5.1 Main Effects

The main effects of the variables followed a distinct pattern. English proficiency (*LexTALE*) was a highly significant positive predictor across all eight models for both pitch-related and timing-related prosody skills. Its 95% bootstrapped confidence intervals consistently excluded zero, confirming its robust predictive power. In contrast, L1 language distance did not show any significant main effects in any of the models, as all of its 95% CIs included zero.

Regarding the specific music aptitude subtests, the findings showed that, for pitch-related prosody skills (*Prosody Pitch Total*), only *PROMS Tuning* (β=0.2, 95% CI: 0.00, 0.39) was a significant positive predictor. This indicates that the ability to perceive musical tuning is positively associated with the perception of pitch-related prosody. The other three subtests, however, did not have significant predictive power. For timing-related prosody skills (*Prosody Time Total*), only *PROMS Tempo* (β=0.23, 95% CI: 0.06, 0.41) was a significant positive predictor. This highlights a specific link between the perception of musical tempo and the perception of prosodic timing.

4.5.2 Interaction Effects

The results showed no statistically significant interaction effects for either L1 language distance or English proficiency with any of the four music aptitude subtests. All of the 95% bootstrapped confidence intervals for the interaction terms (*Predictor × L1* and *Predictor × LexTALE*) contained zero. This indicates that the influence of a learner's music aptitude on their prosody perception (for both pitch and timing) does not significantly vary based on their L1 background or their general English proficiency within this sample. In other words, the relationship between these musical abilities and prosody perception appears to be relatively consistent regardless of these learner-specific factors.

In summary, the analyses for RQ4 show that specific aspects of music aptitude, particularly tuning perception, significantly predicts pitch-related prosody perception. Tempo perception also significantly predicts timing-related prosody. Melody and accent abilities, however, did not show statistically significant predictive power. English proficiency consistently emerged as a powerful, general predictor for both prosodic dimensions. Neither L1 language distance nor English proficiency were found to significantly modulate the impact of any aspects of music aptitude on neither dimension of prosody perception.

4.6 L1 as a Categorical Variable: A Follow-up Study

As previous descriptive analysis have shown, a closer examination of the L1 background data, which was initially quantified as a numerical factor (*L1 Distance*), revealed that this set of data was more akin to distinct categorical groupings rather than a fine-grained continuous spectrum. Specifically, the L1 backgrounds of the participants naturally fell into established rhythmic typologies: stress-timed, syllable-timed, and mixed/mora-timed languages (Abercrombie, 1967; Dauer, 1983; Grabe & Low, 2002; Nespor et al., 2011; Ramus et al., 1999). Therefore, the initial finding of no effects of L1 language background may be due to the numerical representation used in the initial models, which may have masked important group-specific effects. To more appropriately capture these inherent categorical distinctions within the L1 data and to explore their potential moderating roles more accurately, follow-up sets of analyses were performed.

4.6.1 Follow-up Analysis for RQ1 and RQ2

To address RQ1, and since Spearman correlation cannot be used on categorical variables, a follow-up regression model was conducted. This follow-up analysis of RQ1 therefore also answered RQ2, as it simultaneously examined the associations between all key variables (RQ1) and also assessed music aptitude's predictive power while controlling for the influence of L1 background and English proficiency (RQ2). In this model, L1 was treated as a categorical variable, with "stress-timed" serving as the reference group. To ensure the robustness of the findings, bootstrapping with 1000 replicates was employed to generate 95% confidence intervals for the regression coefficients.

Table 4.8

Standardised Regression Coefficients (β) for Regression Model (RQ1&2 Follow-up)

Predictor	Estimate (β)	95% CI
(Intercept)	0.24	(-0.35, 0.78)
PROMS Total	0.24	(0.06, 0.43)*
L1 Category Mixed Mora Timed	-0.1	(-0.78, 0.61)
L1 Category Syllable Timed	-0.31	(-0.93, 0.31)
LexTALE	0.3	(0.12, 0.50)*

Note. A 95% bootstrapped confidence interval that does not contain zero indicates a significant effect, marked with an asterisk (*). Estimates and CI values are rounded to two decimal places for readability. The stress-timed category serves as the reference group for L1 comparisons.

As summarised in Table 4.8, the results were consistent with the initial findings. *PROMS Total* remained a significant positive predictor of total prosody perception, as did English proficiency (*LexTALE*), with their respective bootstrapped confidence intervals not containing zero. The L1 rhythmic typology (*L1 Category Mixed Mora Timed* and *L1 Category Syllable Timed*), however, did not show a significant main effect on prosody perception, as its confidence intervals included zero. These findings suggest that while musical aptitude and language proficiency are key factors in prosody perception, the rhythmic typology of a learner's L1, even when treated as a distinct categorical variable, does not independently explain a significant portion of the variance in overall prosody scores within this sample.

4.6.2 Follow-up Analysis for RQ3

The follow-up results for RQ3 where L1 background was re-coded based on its rhythmic types as a categorical variable are shown in Table 4.9.

Table 4.9

Standardised Regression Coefficients (β) for Moderation Model (RQ3 Follow-up)

Variable	Estimate (β)	95% CI
(Intercept)	0.2	(-0.33, 0.68)
PROMS Total	-0.09	(-0.81, 0.52)
L1 Category Mixed Mora Timed	0.08	(-0.60, 0.78)
L1 Category Syllable Timed	-0.27	(-0.82, 0.35)
LexTALE	0.3	(0.12, 0.50)*
PROMS Total: L1 Category Mixed Mora Timed	0.6	(-0.16, 1.44)
PROMS Total: L1 Category Syllable Timed	0.32	(-0.31, 1.09)
PROMS Total: LexTALE	-0.04	(-0.24, 0.21)

Note. A 95% bootstrapped confidence interval that does not contain zero indicates a significant effect, marked with an asterisk (*). Estimates and CI values are rounded to two decimal places for readability. The stress-timed category serves as the reference group for L1 comparisons.

Consistent with the initial analysis, while the main effect of *LexTALE* remains a significant positive predictor, no significant moderation effects from either *LexTALE* or *L1 Category* was revealed. However, *PROMS Total* was found to be insignificant in this model, which is inconsistent with all the previous findings. Given the strong and consistent predictive power of *LexTALE*, one possible explanation might be that a significant portion of the variance in the DV (*Prosody Total*) has already been accounted for by *LexTALE*. This does not necessarily contradict its importance, but rather highlights its more nuanced role within a model that accounts for other strong linguistic factors and potential moderations.

4.6.3 Follow-up Analysis for RQ4

Similar to the previous RQs and consistent with the initial models, the results of this follow-up analysis indicated that neither the main effect of L1 rhythmic category nor its interaction effects with the four music aptitude subtests reached statistical significance. This finding suggests that, within our sample, the moderating influence of L1 rhythmic typology on the relationship between music aptitude and prosody perception is not a significant factor.

Taken together, the results indicates that L1 background, regardless of how it is represented, does not independently explain a significant amount of variance in prosody perception performance within this sample.

V. Discussion

This study investigated the intricate interplay between music aptitude, L1 background, English proficiency, and English prosody perception among adult ESL/EFL learners. Specifically, this study addressed four key questions: (1) To what extent are music aptitude, L1, and language proficiency associated with prosody perception in a non-native language? (2) Does music aptitude predict prosody perception when controlling for L1 and language proficiency? (3) How do L1 and language proficiency moderate the relationship between musical aptitude and prosody perception? (4) Which aspects of prosody perception are most influenced by music aptitude, and how does this vary by L1 and language proficiency? The core findings consistently revealed a significant positive role of overall music aptitude in facilitating non-native prosody perception, particularly through specific links between music tuning and pitch-related prosody perception, and music tempo and timing-related prosody perception. English proficiency, furthermore, emerged as the strongest and most robust predictor of prosody perception across all analyses. These key relationships were consistently observed and remained significant regardless of the specific analytical models employed, including the inclusion of various covariates and interaction terms, confirming the results' reliability.

This chapter will critically discuss these findings, interpret their significance in relation to the existing literature and the theoretical frameworks outlined in Chapter 2, discuss their theoretical and pedagogical implications, acknowledge the study's limitations, and propose avenues for future research.

5.1 Discussion of Key Findings

5.1.1 Association and Prediction of Non-Native Prosody Perception (RQ1 & RQ2)

5.1.1.1 <u>The Role of Music Aptitude</u>

The findings from RQ1 and RQ2 reveals the factors associated with and predictive of non-native English prosody perception. A key finding is that overall music aptitude (PROMS Total) was significantly associated with non-native prosody perception accuracy, with a small to medium effect size in both the correlation model and the regression model. Importantly, this effect remained significant after controlling for English proficiency and L1 phonological distance, with a small-to-moderate effect size, suggesting that music aptitude emerged as a significant positive predictor of non-native prosody perception accuracy.

The finding is consistent with previous studies, including behavioural (Correia et al., 2022; Wong et al., 2007) studies, electrophysiological studies (Marques et al., 2007; Moreno et al., 2009), and meta-analyses (Jansen et al., 2023; Thompson et al., 2025). However, while those previous studies compared musicians with non-musicians (e.g., Kolinsky et al., 2009; Delogu et al., 2010), these simplistic musician/non-musician dichotomy often failed to disentangle the effects of practice from inherent talent. This study, therefore, marks a methodological advancement, as it employed a continuous measure of music aptitude, Mini-PROMS, which deliberately minimises the influence of acquired musical knowledge, providing more robust evidence for concluding that individuals with higher underlying musical potential are generally more attuned to non-native prosody. This enables the direct investigation into the role of innate music aptitude, which is independent of acquired music experiences. On the other hand, this observed small-to-moderate effect size suggests that while music aptitude plays a role, it may only be one of the several contributing factors, suggesting that L2/FL acquisition is a complex process where multiple factors interplay with each other.

5.1.1.2 The Role of English Proficiency

Beyond music aptitude, English proficiency stood out as the most prominent and consistent positive predictor of non-native prosody perception in the current study, with a moderate effect size and statistical significance across all analytical models. This finding can be interpreted using the Perceptual Assimilation Model (PAM; Best et al., 1994), which suggests that as learners become more proficient, their auditory systems become more “tuned” to the specific phonological and prosodic categories of the target language, thereby reducing the reliance on L1-based perceptual filters. Additionally, this effect may also result from an improved automaticity in perceiving and processing the target language’s suprasegmental features, as enhanced proficiency requires less cognitive resources to capture the prosodic cues, as described by the Automatic Selective Perception (ASP) Model (Strange, 2011).

This finding aligns with previous empirical evidence. Prior work has demonstrated that low proficiency learners perform worse when producing non-native prosodic features, leading some researchers to conclude that "low proficiency precludes near-native production of prosodic features" (Gut, 2003, p. 149). This is extended to prosody perception as well, as inferred from Asano's (2018) study, which found that the less proficient the learners were, the more impaired their performance was under increased task demands. However, Asano's study implicitly measured proficiency as a categorical factor by comparing broad groups, which limits the generalisability of findings and cannot delineate the role of proficiency as spanning a continuous spectrum. Our study, therefore, fills this gap by explicitly measuring proficiency level as a continuous numerical variable through LexTALE, further supporting that higher proficiency predicts better prosody perception in the non-native language.

5.1.1.3 The Role of L1 Distance

In contrast, L1 distance consistently demonstrated no statistically significant association or predictive power regarding non-native prosody perception, irrespective of whether it was treated as a numerical or categorical variable. This result partially challenges the predictions of theoretical models like the PAM (Best et al., 1994), which suggests that L1 phonological distance should directly influence the assimilation of non-native sounds, therefore should be associated with non-native perception. This is also in conflict with prior studies as well, which have also highlighted the impact of L1 background, though mostly on the differences between toned and non-toned languages and how they influence pitch perception (e.g., Alexander et al., 2005; Chan & Leung, 2020). Since prosody encompasses both pitch cues and temporal cues, and this study's target language, English, falls into different tonal and rhythmic categories (stress-timed, syllable-timed, or mixed/mora-timed) than Mandarin or Thai (most common target languages used in prior studies), it is possible that the different finding of this study results from this difference in target language and dimensions of prosody perception tested.

However, an alternative explanation might be due to this study's participant sample's intermediate to upper-intermediate English proficiency level, which may have masked or mitigated the L1 filtering effects seen in novice learners (Best et al., 1994; Jansen et al., 2023). This is further supported by Flege et al.'s Speech Learning Model (SLM; Flege, 1995; Flege et al., 2021). The SLM posits that during the novice stage, initial L2 input is often assimilated to the established L1 phonetic categories, as Best et al.'s (1994) PAM model claimed; but as L2 learners gain increasing exposure in the target language, they begin to form new phonetic categories for L2 sounds that are distinct from their L1, indicating that the reliance on and interference from L1 categories gradually diminishes. Thus, for learners at an intermediate or advanced stage, the direct influence of their L1's phonological distance on the perception of new L2 sounds might be less pronounced than at the very early stages of acquisition, as their developing L2 system takes precedence. This suggests that the absence of a significant L1 distance effect in our study may be explained by the participants' established proficiency, which allowed them to overcome, or at least significantly reduce, L1's influence.

In fact, empirical evidence have demonstrated how a listener's L1 phonological system may shape their perception of non-native prosodic distinctions at the initial stage. For example, Dupoux et al. (1997) tested Spanish speakers (whose L1 uses accent to distinguish words) and French speakers (whose L1 does not) on an accent-based discrimination task using an artificial language, where subjects could be considered low proficiency learners. They found that French subjects experienced significantly more difficulty than Spanish subjects in discriminating accent, and Spanish subjects were unable to ignore task-irrelevant accent differences in a phoneme-based task, whereas French subjects had no such difficulty. This illustrates the strong L1 filtering effect at low proficiency stages, where L1 patterns are automatically imposed on new linguistic input, even when irrelevant, leading to perceptual difficulties for learners whose L1s lack such distinctions. This finding supports the SLM, implying that L1 filtering may only be effective during the initial stage, while at higher proficiency levels, as in our study, learners might have moved beyond this initial L1 filtering and established more L2-specific perceptual abilities.

Additionally, such a non-significant finding may also be due to this study's focus on prosody perception instead of production, which is widely recognised as an area where L1 influence is particularly pronounced, commonly contributing to the phenomenon of a "foreign accent" (e.g., Flege & Fletcher, 1992). A typical example is Gut's (2003) study which found that speakers whose L1 does not possess the same durational differences as in the target language failed to produce these prosodic contrasts. This disparity may be explained by the higher requirements of non-native production, involving more processes like mastering the articulatory gestures (e.g., the vocal organs) required to produce the already perceived and stored L2 sounds, which goes beyond the demands of perception alone (Flege, 1995). Consequently, learners are often capable of perceiving more non-native distinctions than they can produce, which might explain this inconsistency in findings between production-focused studies like Gut (2003) and our perception-focused study.

Another explanation for the current study's non significant findings might be methodological limitations, specifically, the measurement of L1 distance. While the URIEL metric for quantifying L1 distance provided a continuous measure, it proved to be too coarse-grained, as many phonologically distinct L1s were assigned very identical distance scores (e.g., many were assigned the specific distance scores of 0.2804, 0.3433, 0.5687, etc.), thereby reducing its analytical power. Though a follow-up analyses categorising L1 by rhythmic type was conducted to partially address this methodological limitation of quantifying the L1 distance, this follow-up analysis also yielded non-significant results. This is likely due to the imbalanced distribution of L1s across these categories within our sample, where certain types had too few participants for statistically meaningful comparisons, which might result from insufficient statistical power (Type II error) (See section 5.4 for further discussion).

5.1.2 The Moderating Roles of L1 and Language Proficiency (RQ3)

Following the exploration of direct relationships, RQ3 investigated whether L1 background or English proficiency moderate the relationship between music aptitude and prosody perception. The statistical analyses find that neither L1 language distance nor English proficiency moderated the relationship between music aptitude and prosody perception, as the bootstrapped confidence intervals for the interaction terms consistently included zero.

This finding contrasts with Jansen et al. (2023)'s suggestion that proficiency might modulate the music-prosody link across different stages of acquisition. Instead, our results imply that the beneficial effect of music aptitude is relatively consistent across the intermediate to upper-intermediate proficiency spectrum observed in our sample. This might mean that music aptitude provides a foundational, persistent advantage regardless of how proficient a learner becomes within this range.

However, an alternative explanation for this lack of evidence of a moderating effect of language proficiency may result from this unbalanced proficiency level distribution. Given that the majority of participants possess intermediate to high proficiency levels of English (LexTALE mean of 48.8/60), the moderating effect of proficiency might be less pronounced than it would be among less proficient learners, as they are already relatively more attuned to the prosodic features of the target language, as suggested by the PAM (Best et al., 1994) and ASP model (Strange, 2011).

In fact, a possible explanation might be that the role of proficiency may not operate as a continuous moderator across the entire spectrum, but rather as a "threshold" that enables cross-domain transfer to become effective. This can be inferred from a previous study by Wiener and Bradley (2023), which employed an eight-week intervention to Mandarin beginners either in L2 Mandarin learning (+L2), in music training (+Music), or both (+L2+Music). The results showed that only the +L2+Music group demonstrated a significant increase in sensitivity to Mandarin lexical contrasts, while the +Music group showed only marginal improvement and the +L2 group exhibited no statistically significant gains.

This interesting pattern indicates that music training alone, without concurrent L2 instruction that develops proficiency, may not significantly improve L2 prosody perception sensitivity. However, when music training is accompanied by L2 teaching that enhances language proficiency, it improves L2 prosodic sensitivity. This suggests that the improved L2 proficiency might have created a necessary proficiency "threshold" for musical abilities to effectively contribute to L2 prosodic perception. Without this developed threshold proficiency in the L2, music training appears to have limited impact on L2 prosody perception. Though not providing direct strong evidence for a threshold effect, Wiener and Bradley's finding tentatively suggests that the moderating role of target language proficiency may primarily manifest among less proficient learners, a hypothesis which needs to be verified by future studies.

As for L1 background, the consistent lack of moderation suggests that, for this sample and with the measures used, L1 background does not directly modulate how music aptitude impacts non-native prosody perception. One possible explanation for this is a genuine absence of such a moderation effect, as supported by previous empirical evidence. For example, Choi (2022) found that musical ability consistently facilitates English stress perception across Cantonese and English listeners, while the interaction between musical ability and language was not significant, consistent again L1 background not moderating the music-prosody relationship. This might be because learners from phonologically distant L1s employ some kind of compensatory strategies, perhaps drawing more heavily on their general auditory processing skills like music aptitude to overcome L1 interference, thus masking a direct negative effect of L1 distance at the group level.

However, this lack of moderation effect could also be attributed to insufficient statistical power (Type II error) to detect subtle interactions, or as previously mentioned, a limitation due to the large-grained quantification of L1 distance and the unequal distribution of languages within the sample. This is partially supported by the visual inspection of the moderation plot for L1 background, which demonstrates a non-significant trend where the positive predictive power of music aptitude on prosody perception appears slightly stronger when a learner's L1 is phonologically closer to the target language. In other words, this suggests that, with larger statistical power, a more fine-grained metric for L1 distance, or a more balanced and widespread distribution of participants' L1s, there might be a trend that the further the L1 is from English (the target language), the more music aptitude predicts English prosody perception.

5.1.3 Specific Aspects of Music Aptitude and Prosody Perception (RQ4)

RQ4 aimed to uncover which specific aspects of music aptitude (melody, tuning, accent, or tempo) most influenced specific dimensions of prosody perception (pitch-related vs. timing-related), and how these relationships might be moderated.

The findings suggest a specificity of transfer: tuning perception in the music aptitude test (PROMS Tuning) was a significant positive predictor of pitch-related prosody skills (Prosody Pitch Total), and tempo perception in the music aptitude test (PROMS Tempo) was a significant positive predictor of timing-related prosody skills (Prosody Time Total). Conversely, PROMS Melody and PROMS Accent were not statistically significant predictors for either prosody dimension. It is crucial to note that such a non-significant finding indicates a lack of statistically detectable effect within the current study, rather than definitive proof of an absence of effect. While it is important not to over-interpret non-significant results, this observed pattern is consistent with a highly specific cross-domain transfer effect, where the ability to precisely discriminate musical pitch (tuning) directly translates to the ability to perceive pitch variations in English prosody (e.g., intonation, focus pitch) while sensitivity to musical tempo enhances the perception of temporal aspects of English prosody (e.g., duration, pauses, rhythm). This echoes a number of previous findings. For example, speakers of languages that use durational cues to distinguish meaning (e.g., Finnish, where words can be distinguished from vowel duration) were found to be more accurate at detecting durational variations in non-speech sounds (Marie et al., 2012), and speakers of tone language were found to be better at discriminating sung intervals (e.g., Pfordresher and Brown, 2009).

A more fine-grained study on pitch-related music-to-language transfer was conducted by Ong et al. (2016), who trained adults to acquire pitch-based elements either in Thai lexical tone minimal pairs or musical chord minimal pairs, finding that learners showed improvements in discrimination of test items in the domain other than what they were trained on, which suggests a cross-domain transfer. However, while Ong et al. (2016) observed this cross-domain benefit specifically for pitch, their study's design limits the generalisability of their claim regarding dimension-specific transfer. Their research exclusively focused on pitch-related stimuli and did not investigate other prosodic dimensions (e.g., timing) or overall musical ability. Consequently, the pitch-specific gains they observed might merely reflect a broader, undifferentiated music-to-language transfer, rather than providing robust evidence for the precise, domain-specific transfer that they claimed. The current study, therefore, fills this gap as its multi-dimensional measures of both music aptitude and prosody more robustly suggest such a domain-specific relationship, though this cross-sectional design cannot soundly indicate a causal transfer effect.

The non-significant predictive power of PROMS Melody and Accent in this study contrasts with findings from similar research. For instance, Li et al.’s (2024) study used PROMS to investigate the specific influences of various musical dimensions on speech imitation skills across six typologically diverse languages (excluding English) among Chinese and Catalan speaking participants. They reported that melody was the only significant predictor among Catalan L1 speakers, and accent was the only significant predictor among Chinese speakers. However, their finding possessed a methodological flaw, as they did not apply a correction for multiple comparisons across the four PROMS subtests. This omission suggests that the reported significant finding for accent among Chinese speakers might not hold up under stricter statistical scrutiny. Therefore, a more robust interpretation of Li et al.’s study, when considering a corrected significance threshold, is that only melody significantly predicted general phonetic production skills among Catalan (an intonation language) speakers. The authors attributed this to Catalan speakers benefiting from the transfer of better musical melody perception skills into enhanced abilities for detecting fine-grained intonational differences at the sentence level, an area where Catalan speakers often face challenges.

Our study’s differing results, where neither PROMS Melody nor Accent significantly predicted English prosody perception, may thus be explained by the typological differences of the target language (English) and the nature of the prosodic perception task. English is a stress-timed language (Abercrombie, 1967), and while pitch variation is crucial for English prosody (e.g., for intonation and focus), the melodic contour in English prosody might not be as finely discriminative for meaning as it is in tone languages (like Mandarin, which Li et al. also investigated). Similarly, while accent (rhythm) is important, the specific aspects measured by PROMS Accent may not directly align with the rhythmic properties most salient or challenging for non-native perception of English stress timing. Furthermore, it is also possible that, since the stimuli in our prosody perception test were specifically designed to be more accessible for non-native speakers (by using more distinct 0% vs. 100% pitch and duration contrasts), it did not sufficiently challenge participants in ways that would highlight subtle melodic or accent perception differences from music. This could have led to a ceiling effect for some participants, as indicated by the left-skewed distribution of prosody perception scores, thereby obscuring potential associations with PROMS Melody and Accent.

Additionally, in regard of the moderation effect of L1 distance and English proficiency, the findings were consistent with RQ3. This again suggests that the observed specific transfers might be relatively consistent effects (if verified that such an absence in moderation effects is not due to lack of statistical power). This suggests a possible link between music and language perception at a more fundamental auditory-perceptual level.

5.2 Theoretical Implications

There are several important implications of the findings of this study, which advances our understanding of the music-language relationship and individual differences in L2/FL acquisition. Firstly, the finding that there is a direct link between inherent musical potential and non-native prosody perception aligns well with the OPERA hypothesis (Patel, 2011), suggesting that the "O" (Overlapping neural processing) and "P" (Precision for fine-grained acoustic distinctions) components may be rooted in fundamental, possibly innate, auditory processing capacities for music. While previous research supporting OPERA has largely emphasised the role of explicit music training in facilitating language acquisition, our findings extend this framework by providing additional evidence for the contribution of inherent music aptitude, broadening our understanding of pathways through which musical abilities can influence L2/FL perception.

Secondly, the discovery of specific links between musical tuning perception and pitch-related prosody, and musical tempo perception and timing-related prosody, provides initial evidence of cross-domain transfer, and tentatively, that this transfer is domain-specific. This goes beyond broad statements of music-language connections and supports a more modular understanding of shared perceptual mechanisms. It suggests that the auditory system is not just generally sensitive to acoustic cues but has specialised processing mechanisms that can potentially transfer across domains.

Thirdly, the consistent finding that L1 background (whether regarded as continuous distance or categorical rhythmic typology) did not significantly predict or moderate non-native prosody perception somehow challenges existing theoretical models like the PAM (Best et al., 1994) and ASP Model (Strange, 2011), which emphasises L1 as a filter for L2 sounds. This study tentatively suggests that for the specific context of adult intermediate to upper-intermediate English learners and prosody perception as measured here, the influence of L1 might be less direct than often assumed. It prompts a re-evaluation of the exact contexts, stages, or linguistic features that L1 may consistently exert influence, especially when compensatory mechanisms like music aptitude play a role.

5.3 Pedagogical Implications

Firstly, the robust predictive power of general English proficiency on prosody perception suggests that, in L2 or FL teaching of prosody, while specific interventions might be beneficial, a highly impactful pedagogical strategy could involve prioritising holistic development of overall language competence. Teachers might consider continuing to emphasise broad exposure to the target language and extensive engagement with authentic materials that naturally expose learners to varied prosodic patterns. As proficiency increases, learners' implicit understanding of the phonological and prosodic system could likely improve as well, potentially facilitating more automatic and accurate perception.

Secondly, the identification of specific links between musical tuning and pitch-related prosody, and musical tempo and timing-related prosody, offers potential suggestions for targeted pedagogical interventions, though this should be further confirmed through controlled intervention studies. A tentative suggestion at this point is that, for learners struggling with specific prosodic aspects, teachers might consider incorporating targeted exercises designed to enhance relevant perceptual skills. For instance, ear-training exercises focusing on pitch discrimination (identifying rising/falling intonation, pitch accents) could indirectly aid pitch-related prosody. Similarly, activities emphasising rhythmic patterns, beat alignment, or maintaining a steady tempo might support the perception of duration and stress timing in the target language.

More broadly, although this cross-sectional study cannot provide any causal evidence as to music training's effects on language learning, this study's findings do offer insights into individual difference, suggesting educators to consider a learner's music aptitude as a potential factor contributing to individual variation in L2/FL perception. While aptitude is a relatively stable trait and not directly amenable to instruction, understanding these individual differences could enable teachers to develop more personalised learning strategies. For instance, a learner with lower musical tempo aptitude who struggles with English rhythm might benefit from more explicit exposure to rhythmic drills, or a higher degree of visual support for temporal cues. Similarly, a learner with high tuning aptitude might quickly grasp intonation patterns, allowing the teacher to focus on other areas. In general, this study's findings suggest that teaching instructions might be tailored to address specific perceptual challenges in L2/FL learning as reflected by different aspects of individual differences.

5.4 Limitations and Future Research

Despite its contributions, this study still has several limitations that warrant consideration when interpreting its findings and informing future research.

First and foremost, the study employed a cross-sectional design, which, by its nature, cannot infer causality. While associations and predictive relationships were observed, it is not sufficient to claim that higher music aptitude causes better prosody perception. Future intervention-based studies should incorporate rigorously designed music training interventions to establish such causal links.

Secondly, the sample size (n=100) in this study presented a significant limitation, particularly for detecting moderation effects. As indicated by the G*Power analysis in Chapter 3, the power analysis for interaction terms in RQ3 required a substantially larger sample of approximately 133 participants to detect a small effect size ($f^2 = 0.06$) with 80% power. While the study's sample size was sufficient for detecting main effects and correlations (requiring 55-84 participants), it was slightly underpowered for interactions. Therefore, the non-significant moderation effects found for L1 background and English proficiency on the music-prosody relationship (RQ3 and RQ4) should be interpreted with caution, as it is possible that genuine moderation effects exist but were not detected due to insufficient statistical power. Future studies, therefore, should aim for significantly larger sample sizes (e.g., 400+ participants) to robustly test the moderating roles of L1 background and L2 proficiency, as well as any other potential confounding factors. Additionally, the sample was predominantly composed of "music loving non-musicians" or "non-musicians", with intermediate to upper-intermediate English proficiency, which limits the generalisability of findings to other participant demography, another issue that future studies should address.

Thirdly, while the chosen measures (Mini-PROMS, MBOPP, LexTALE) were selected for their psychometric properties and suitability for the research questions, they also present specific limitations. LexTALE, though provides a robust indicator of general English proficiency, is vocabulary-based, thus cannot capture the full spectrum of linguistic competence (e.g., grammar, pragmatics, fluency). Nevertheless, the consistent and significantly positive predictive power of such a lexical-based (thus irrelevant to prosody) test on prosody perception is particularly compelling. This finding more firmly suggests that even a foundational component of lexical knowledge is intertwined with and predictive of prosodic processing, which reinforces the broader argument that general, holistic language competence plays a crucial role in the development of L2/FL perception skills.

Continuing the discussion on methodological considerations, other test materials also present certain limitations. The prosody perception test, MBOPP, while excellent for assessing focus and phrase boundary perception across pitch and timing, does not cover all aspects of English prosody (e.g., emotional prosody, discourse functions). This highlights an avenue for future work to explore whether music abilities or other individual differences are associated with or predictive of these broader prosody dimensions. Furthermore, the adaptation of MBOPP stimuli to be more accessible for non-native speakers (as discussed above in section 3.2.2) might have led to a ceiling effect for some advanced learners, as shown in the left skew of scoring distribution. This potentially limits the strength of correlations with other variables. Future studies, therefore, should aim to address these methodological limitations by implementing more targeted and comprehensive designs. This could involve utilising stimuli with subtler prosodic contrasts to avoid ceiling effects, especially for higher proficiency learners. Achieving this fine-grained alignment with participants' proficiency level might necessitate researchers to have an overview of participant profile prior to the main experiment, possibly by collecting their self-reported proficiency levels or conducting a pre-test on proficiency. An alternative might be to recruit participants across a wider spectrum of proficiency levels to fully explore how proficiency interacts with music abilities in prosody perception.

Additionally, the operationalisation of L1 distance presents many methodological limitations. The URIEL L1 distance metric used to quantify L1 distance, though ideally effective for providing a continuous measure of phonological distance from English, proved to be too coarse-grained for analysis. Specifically, a large number of phonologically distinct L1s were assigned very similar distance scores (e.g., clustering around certain numbers like 0.2804, 0.3433, 0.5687), thus reducing its analytical power and requires a more precise metric to capture any potential effects of L1 background in future studies. Furthermore, the categorical classification of L1s by rhythmic type (stress-timed, syllable-timed, mixed/mora-timed) suffered from an imbalanced distribution across categories within our sample, which limited this study's ability to draw robust conclusions about the moderating influence of specific L1 rhythmic characteristics on prosody perception, as certain categories had too few participants to yield statistically meaningful comparisons. Future studies could include larger sample sizes for each type of L1s, and ideally include a wider range of L1s with more distinct prosodic characteristics relative to English (e.g., more tonal languages, languages with very different rhythmic patterns) to explore L1 transfer effects more comprehensively. The target language could extend to more less-studied languages as well, as most studies that have looked at music aptitude and prosody have looked at English, which questions whether current findings may be language specific.

VI. Conclusion

This study investigated the complex relationships between music aptitude, L1 background, English proficiency, and non-native English prosody perception. The findings suggest a positive role for overall music aptitude in facilitating non-native prosody perception, especially through specific links between musical tuning and pitch-related prosody, and musical tempo and timing-related prosody. English proficiency emerged as the strongest and most consistent predictor to prosody perception, underscoring the importance of general language development in L2/FL perception skills. L1 background, however, did not show significant direct or moderating effects, which may reflect the mitigating influence of higher proficiency in the target language among the current sample, or may be due to limitations in statistical power and possibly in the measures used in the current study.

This research carries significant theoretical and pedagogical implications. It confirms the OPERA hypothesis and extends its scope to innate music aptitude, thus advancing the understanding of shared auditory mechanisms. It also contributes significantly to the fields of second language acquisition and psycholinguistics by advancing our understanding of an important area of individual differences in L2/FL learning. Pedagogically, the findings could be interpreted as suggesting a focus on holistic proficiency development while also indicating the potential for targeted auditory training tailored to specific prosodic challenges. Despite some limitations, this study provides valuable empirical evidence, paving the way for future research to further unveil the intricate relationships between music and language.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Interspeech 2005* (pp. 397–400). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2005-271

Asano, Y. (2018). Discriminating non-native segmental length contrasts under increased task demands. *Language and Speech*, *61*(3), 409–429. https://doi.org/10.1177/0023830917731907

Bababekov, Y. J., Hung, Y. C., Hsu, Y. T., Udelsman, B. V., Mueller, J. L., Lin, H. Y., & Chang, D. C. (2019). Is the power threshold of 0.8 applicable to surgical science?—Empowering the underpowered study. *The Journal of Surgical Research*, *241*, 235–239. https://doi.org/10.1016/j.jss.2019.03.062

Barbaroux, M., Norena, A., Rasamimanana, M., Castet, E., & Besson, M. (2021). From psychoacoustics to brain waves: A longitudinal approach to novel word learning. *Journal of Cognitive Neuroscience*, *33*(1), 8–27. https://doi.org/10.1162/jocn_a_01629

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). The MIT Press.

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education*, *7*(2), 113–123. https://doi.org/10.1111/mbe.12016

Bodner, E., Aharoni, R., & Iancu, I. (2012). The effect of training with music on happiness recognition in social anxiety disorder. *Journal of Psychopathology and Behavioral Assessment*, *34*(4), 458–466. https://doi.org/10.1007/s10862-012-9304-7

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: Language and Cognition*, *19*(5), 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*(4), 774–808. https://doi.org/10.1111/lang.12159

Bradley, E. D. (2012). *Crosslinguistic perception of pitch in language and music* [Unpublished doctoral dissertation]. University of Delaware.

Bradley, E. D. (2018). A comparison of stimulus variability in lexical tone and melody perception. *Psychological Reports*, *121*(4), 600–614. https://doi.org/10.1177/0033294117734832

Brown, A. W. (1928). The reliability and validity of the Seashore Tests of Musical Talent. *Journal of Applied Psychology*, *12*(5), 468–476. https://doi.org/10.1037/h0072753

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248. https://doi.org/10.1177/0305735611425896

Chan, R. K. W., & Leung, J. H. C. (2020). Why are lexical tones difficult to learn? Insights from the incidental learning of tone-segment connections. *Studies in Second Language Acquisition*, *42*(1), 33–59. https://doi.org/10.1017/s0272263119000482

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience*, *31*(6), 751–760. https://doi.org/10.1080/23273798.2016.1156715

Chenoweth, V. (1980). *Melodic perception and analysis: A manual on ethnic melody*. Summer Institute of Linguistics.

Choi, W. (2020). The selectivity of musical advantage: Musicians exhibit perceptual advantage for some but not all Cantonese tones. *Music Perception*, *37*(5), 423–434. https://doi.org/10.1525/mp.2020.37.5.423

Choi, W. (2022). What is 'music' in music-to-language transfer? Musical ability but not musicianship supports Cantonese listeners' English stress perception. *Journal of Speech, Language, and Hearing Research*, *65*(10), 3894–3909. https://doi.org/10.1044/2022_JSLHR-22-00175

Choi, W., Ling, C. L. K., & Wu, C. H. J. (2024). Musical advantage in lexical tone perception hinges on musical instrument: A comparison between pitched musicians, unpitched musicians, and nonmusicians. *Music Perception*, *41*(5), 360–377. https://doi.org/10.1525/mp.2024.41.5.360

Cogo-Moreira, H., de Avila, C. R. B., Ploubidis, G. B., & de Jesus Mari, J. (2013). Effectiveness of music education for the improvement of reading skills and academic achievement in young poor readers: A pragmatic cluster-randomized, controlled clinical trial. *PLoS ONE*, *8*(3), Article e59984. https://doi.org/10.1371/journal.pone.0059984

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cole, J. (2015). Prosody in context: a review. *Language, Cognition and Neuroscience*, *30*(1-2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America*, *131*(6), 4756–4769. https://doi.org/10.1121/1.4714355

Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2022). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*, *22*(5), 894–906. https://doi.org/10.1037/emo0000770

Cui, A., & Kuang, J. (2019). The effects of musicality and language background on cue integration in pitch perception. *The Journal of the Acoustical Society of America*, *146*(6), 4086–4096. https://doi.org/10.1121/1.5134442

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62.

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schon, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research*, *1455*, 75–89. https://doi.org/10.1016/j.brainres.2012.03.034

Delogu, F., Lampis, G., & Belardinelli, M. (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

Dittinger, E., Barbaroux, M., d'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, *28*(10), 1584–1602. https://doi.org/10.1162/jocn_a_00997

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*(3), 406–421. https://doi.org/10.1006/jmla.1996.2488

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Faßhauer, C., Frese, A., & Evers, S. (2015). Musical ability is associated with enhanced auditory and visual cognitive processing. *BMC Neuroscience*, *16*, 59. https://doi.org/10.1186/s12868-015-0200-4

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, *37*(1), 1–9. https://doi.org/10.3758/MC.37.1.1

Field, A., Miles, J., Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, *91*(1), 370–389. https://doi.org/10.1121/1.402636

Flege, J. E., Aoyama, K., & Bohn, O. S. (2021). The revised speech learning model (SLM-r) applied. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 84–118). John Benjamins Publishing Company.

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS ONE*, *10*(9), Article e0138715. https://doi.org/10.1371/journal.pone.0138715

Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249. https://doi.org/10.1016/j.tics.2006.04.002

Fuller, C. D., Galvin, J. J., 3rd, Maat, B., Free, R. H., & Baßskent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations?. *Frontiers in Neuroscience*, *8*, 179. https://doi.org/10.3389/fnins.2014.00179

Gagne, F. (1999). Nature or nurture? A re-examination of Sloboda and Howe's (1991) interview study on talent development in music. *Psychology of Music*, *27*(1), 38–51. https://doi.org/10.1177/0305735699271004

Gordon, E. (1965). The Musical Aptitude Profile: A new and unique musical aptitude test battery. *Bulletin of the Council for Research in Music Education*, (6), 12–16.

Gordon, E. (1989). *Advanced measures of music audiation*. Gia Publications.

Götz, A., Liu, L., Nash, B., & Burnham, D. (2023). Does musicality assist foreign language learning? Perception and production of Thai vowels, consonants and lexical tones by musicians and non-musicians. *Brain Sciences*, *13*(5), 810. https://doi.org/10.3390/brainsci13050810

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 377–401). Mouton de Gruyter.

Gut, U. (2003). Prosody in second language speech production: The role of the native language. *Fremdsprachen Lehren und Lernen*, *32*, 132–152.

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology*, *6*, 1962. https://doi.org/10.3389/fpsyg.2015.01962

Hirst, D. (2004). The phonology and phonetics of speech prosody: between acoustics and interpretation. In J. C. L. Vossen, M. J. Spivey, & K. McRae (Eds.), *Proceedings of the 2nd International Conference on Speech Prosody* (pp. 67–70). International Speech Communication Association.

Houghton, D. (2021). *A Influência da Aprendizagem Musical no Desempenho na MBEA* (Publication No. 30846493) [Doctoral dissertation, Universidade Federal de Minas Gerais]. ProQuest Dissertations & Theses Global.

Jansen, N., Harding, E. E., Loerts, H., Başkent, D., & Lowie, W. (2023). The relation between musical abilities and speech prosody perception: A meta-analysis. *Journal of Phonetics*, *101*, Article 101278. https://doi.org/10.1016/j.wocn.2023.101278

Jasmin, K., Dick, F., & Tierney, A. T. (2021). The Multidimensional Battery of Prosody Perception (MBOPP) (Version 2). *Wellcome Open Research*. https://doi.org/10.12688/wellcomeopenres.15607.2

Jasmin, K., Dick, F., Holt, L., & Tierney, A. (2018). Tailored perception: listeners' strategies for perceiving speech fit their individual perceptual abilities. *bioRxiv*. https://doi.org/10.1101/263079

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

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

Kim, H., Yoon, T., Cole, J., & Hasegawa-Johnson, M. (2006). Acoustic differentiation of L- and LL% in switchboard and radio news speech. In *Proceedings of Speech Prosody 2006*. Dresden, Germany.

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language network" serves the processing of music. *NeuroImage*, *17*(2), 956–966. https://doi.org/10.1006/nimg.2002.1154

Koelsch, S., Gunter, T. C., Wittforth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577. https://doi.org/10.1162/089892905774597290

Kolinsky, R., Cuvelier, H., Goetry, V., Peretz, I., & Morais, J. (2009). Music training facilitates lexical stress processing. *Music Perception*, *26*(3), 235–246. https://doi.org/10.1525/mp.2009.26.3.235

Kragness, H. E., Swaminathan, S., Cirelli, L. K., & Schellenberg, E. G. (2020). Individual differences in musical ability are stable over time in childhood. *Developmental Science*, *24*(2), Article e13081. https://doi.org/10.1111/desc.13081

Kunert, R., Willems, R. M., & Hagoort, P. (2016). An independent psychometric evaluation of the PROMS measure of music perception skills. *PLoS ONE*, *11*(7), Article e0159103. https://doi.org/10.1371/journal.pone.0159103

Ladefoged, P. (1975). *A course in phonetics*. Harcourt Brace Jovanovich.

Langus, A., Marchetto, E., Bion, R., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*, *66*(1), 285–306. https://doi.org/10.1016/j.jml.2011.09.004

Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PLoS ONE*, *7*(12), Article e52508. https://doi.org/10.1371/journal.pone.0052508

Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235–3248. https://doi.org/10.1121/1.2990713

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, *53*(1), 380.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A valid and efficient measure of receptive vocabulary size in English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0143-6

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

Littell, P., Mortensen, D., Lin, K., Kairis, K., Turner, C., & Levin, L. (2017). URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 8–14). Association for Computational Linguistics.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, *4*(7), 540–545. https://doi.org/10.1038/88444

Maggu, A. R., Wong, P. C. M., Antoniou, M., Bones, O., Liu, H., & Wong, F. C. K. (2018). Effects of combination of linguistic and musical pitch experience on subcortical pitch encoding. *Journal of Neurolinguistics*, *47*, 145–155. https://doi.org/10.1016/j.jneuroling.2018.05.003

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*(10), 2701–2715. https://doi.org/10.1162/jocn.2010.21585

Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, *23*(2), 294–305. https://doi.org/10.1162/jocn.2010.21413

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*(9), 1453–1463. https://doi.org/10.1162/jocn.2007.19.9.1453

Martínez-Castilla, P., Campos, R., & Sotillo, M. (2019). Enhanced linguistic prosodic skills in musically trained individuals with Williams syndrome. *Language and Cognition*, *11*(3), 455–478. https://doi.org/10.1017/langcog.2019.27

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor to language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Menn, L., & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, *25*(4), 341–383. https://doi.org/10.1177/002383098202500403

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, *19*(3), 712–723. https://doi.org/10.1093/cercor/bhn120

Morrill, T. H., McAuley, J. D., Dilley, L. C., & Hambrick, D. Z. (2015). Individual differences in the perception of melodic contours and pitch-accent timing in speech: domain- Support for generality of pitch processing. *Journal of Experimental Psychology: General*, *144*(4), 730–736. https://doi.org/10.1037/xge0000081

Mualem, O., & Lavidor, M. (2015). Music education intervention improves vocal emotion recognition. *International Journal of Music Education*, *33*(4), 413–425. https://doi.org/10.1177/0255761415584292

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), Article e89642. https://doi.org/10.1371/journal.pone.0089642

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, *241*(1-2), 34–42. https://doi.org/10.1016/j.heares.2008.04.013

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J. D., & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, *115*(28), 6630–6639. https://doi.org/10.1073/pnas.1808412115

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain*, *133*(9), 2635–2642. https://doi.org/10.1093/brain/awq178

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 2, pp. 1148–1159). Blackwell Publishing. https://doi.org/10.1002/9781444335262.wbctp0048

Obergfell, A. L., Schmidt, B. M., Stenneken, P., Wittemann, S. K., & Schabmann, A. (2021). Prosodic sensitivity and reading fluency of musicians and non-musicians. *Reading and Writing*, *34*(4), 887–909. https://doi.org/10.1007/s11145-020-10096-4

Ong, J. H., Burnham, D., Stevens, C. J., & Escudero, P. (2016). Naïve learners show cross-domain transfer after distributional learning: The case of lexical and musical pitch. *Frontiers in Psychology*, *7*, 1189. https://doi.org/10.3389/fpsyg.2016.01189

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A. D. (2007). *Music, language, and the brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195123753.001.0001

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

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(5), 717–733. https://doi.org/10.1162/089892998563121

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, *25*(4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357

Peppé, S. J. E. (1998). *Investigating linguistic prosodic ability in adult speakers of English* [Unpublished doctoral dissertation]. University College London.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, *17*(4–5), 345–354. https://doi.org/10.1080/0269920031000079994

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*(1), 58–75. https://doi.org/10.1196/annals.1284.006

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience*, *7*, 30. https://doi.org/10.3389/fnsys.2013.00030

Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., & Dilley, L. C. (2013). Evidence for shared cognitive processing of pitch in music and language. *PLoS ONE*, *8*(8), Article e73372. https://doi.org/10.1371/journal.pone.0073372

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics, 71*(6), 1385-1398.

Plonsky, L., & Oswald, F. L. (2014). How to report and interpret effect sizes in SLA research: A practical guide for primary and meta-analytic research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12078

Qin, Z., Zhang, C., & Wang, W.-S.-Y. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America*, *149*(1), 435–446. https://doi.org/10.1121/10.0003330

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Remijsen, B., & van Heuven, V. J. (1999). Gradient and categorical pitch dimensions in Dutch: Diagnostic test. In *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 2, pp. 1865–1868).

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*, *138*(1), 1–10. https://doi.org/10.1016/j.actpsy.2011.03.007

Schneider, K., Dogil, G., & Möbius, B. (2009). German boundary tones show categorical perception and a perceptual magnet effect when presented in different contexts. In *Interspeech 2009* (pp. 2519–2522). International Speech Communication Association.

Schneider, K., & Lintfert, B. (2003). Categorical perception of boundary tones in German. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 631–634).

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172.x

Schön, D., & Morillon, B. (2018). Music and language. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford handbook of music and the brain* (pp. 391–416). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198804123.013.16

Shockley, J. (2022). *The effects of musical experience and aptitude on phonological skills in a foreign language* [Bachelor's honors thesis]. Harding University.

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

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, *16*(2), 374–381. https://doi.org/10.3758/16.2.374

Sloboda, J. A., & Howe, M. J. A. (1991). Biographical precursors of musical excellence: An interview study. *Psychology of Music*, *19*(1), 3–21. https://doi.org/10.1177/0305735691191001

Speer, S., & Blodgett, A. (2006). Prosody. In M. J. Spivey, M. J. Joanisse, & K. McRae (Eds.), *Handbook of psycholinguistics* (pp. 505–537). Academic Press. https://doi.org/10.1016/B978-012369374-7/50014-6

Stepanov, A., Pavlič, M., Stateva, P., & Reboul, A. (2018). Children's early bilingualism and musical training influence prosodic discrimination of sentences in an unknown language. *The Journal of the Acoustical Society of America*, *143*(1), EL1–EL7. https://doi.org/10.1121/1.5019700

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

Tao, J., Shao, Q., & Gao, X. (2017). Ethics-related practices in Internet-based applied linguistics research. *Applied Linguistics Review*, *8*(4), 321–353. https://doi.org/10.1515/applirev-2016-2024

Thompson, R. M., Salig, L. K., & Slevc, L. R. (2025). Is musical ability related to second-language acquisition? A meta-analysis. *Royal Society Open Science*, *12*(1), 241193. https://doi.org/10.1098/rsos.241193

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech: Effects of music lessons. *Annals of the New York Academy of Sciences*, *999*(1), 530–532. https://doi.org/10.1196/annals.1284.067

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Toh, X. R., Tan, S. H., Wong, G., Lau, F., & Wong, F. C. (2023). Enduring musician advantage among former musicians in prosodic pitch perception. *Scientific Reports*, *13*(1), Article 2657. https://doi.org/10.1038/s41598-023-29733-3

Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the native language on the learning of fundamental frequency in second-language speech segmentation. *Frontiers in Psychology*, *7*, 985. https://doi.org/10.3389/fpsyg.2016.00985

Wagner, M., & Watson, D. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905–945. https://doi.org/10.1080/01690961003589492

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test: A new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188–196. https://doi.org/10.1016/j.lindif.2010.02.004

Wang, X., & Peng, G. (2014). Phonological processing in Mandarin speakers with congenital amusia. *The Journal of the Acoustical Society of America*, *136*(6), 3360–3370. https://doi.org/10.1121/1.4900559

Weidema, J. L., Roncaglia-Denissen, M., & Honing, H. (2016). Top-down modulation on the perception and categorization of identical pitch contours in speech and music. *Frontiers in Psychology*, *7*, 817. https://doi.org/10.3389/fpsyg.2016.00817

Wiener, S., & Bradley, E. D. (2020). Harnessing the musician advantage: Short-term musical training affects non-native cue weighting of linguistic pitch. *Language Teaching Research*, *27*(4), 1016–1031. https://doi.org/10.1177/1362168820971791

Wing, H. (1968). *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity* (2nd ed.). Cambridge University Press.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585. https://doi.org/10.1017/S0142716407070312

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422. https://doi.org/10.1038/nn1872

Yashaswini, L., & Maruthy, S. (2020). Effect of music training on categorical perception of speech and music. *Journal of Audiology & Otology*, *24*(3), 140–148. https://doi.org/10.7874/jao.2019.00500

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*, *1424*(1), 133–144. https://doi.org/10.1111/nyas.13410

Zhang, Y., Geng, T., & Zhang, J. (2018). Emotional prosody perception in Mandarin-speaking congenital amusics. In *Interspeech 2018* (pp. 2196–2200). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2018-91

Appendix A - CUREC Approval

Education (Educ) DREC

15 Norham Gardens, Oxford, OX2 6PY

Applicant: Xinyi Pei

Principal Investigator: Elizabeth Wonnacott

Department: Education

Study title: The Impact of Music Aptitude and L1 on Non-native Language Prosody Perception

(Version: 2.0)

Ethics reference: Education (Educ) DREC - 1451961

Dear Elizabeth Wonnacott,

On behalf of the Committee, I confirm that the above research study described in the application and other supporting documentation submitted to the committee has been carefully considered on behalf of the Education (Educ) DREC in accordance with the University's regulations and policy for ethics approval of research involving human participants, human tissue and/or personal data. The opinion is as follows:

Opinion of Research Ethics Committee: Favourable Opinion

Subject to the following conditions:

Decision Date: 23 Jun 2025, 10:36

Opinion End Date: 6 Nov 2026

If favourable, insurance-provided indemnity arrangements will be in place between the decision date and opinion end date and you may now commence your study activities. Should you plan to continue the research beyond the end date above, it is your responsibility to ensure that you request, and receive, an extension (via amendment) from the committee for indemnity to remain in place. You may be required to provide a justification.

Please note the following:

Amendments: Should there be any subsequent changes to the reviewed study, applications for amendments can be made via the Oxford Ethics Application System (Worktribe Ethics).

Reports: Studies considered by OxTREC are expected to submit an *annual progress report* on each anniversary of study approval, until the study is completed. An end of study report is also required.

Audit: This study may be selected for audit at the discretion of the Research Governance, Ethics and Assurance Team.

Data safety: It is the responsibility of the PI to ensure that all data collected during the course of the study is stored and transferred safely and securely in accordance with University requirements. Further guidance and advice are available from the Research Data Team. Additional information is available at https://researchsupport.web.ox.ac.uk/governance/ethics

Yours Sincerely

Education Ethics Officer

Appendix B - Consent Form and Information Sheet

DEPARTMENT OF EDUCATION
Principle Investigator: Elizabeth Wonnacott
University e-mail:
xinyi.pei@sjc.ox.ac.uk,
elizabeth.wonnacott@education.ox.ac.uk

The Impact of Music Aptitude and L1 on Non-native Language Prosody Perception

CUREC Approval Reference:

General Information

The aim of this research is to investigate how music ability and first language (L1) background influence how people perceive prosody—the rhythm, stress, and intonation of speech—in a non-native language (like English). Previous research suggests that people with stronger musical skills may be better at understanding these speech patterns. However, it is still unclear why this is the case. This study focuses on how different aspects of musical ability, such as melody and rhythm perception, relate to understanding prosodic features like pitch and timing in English. It also looks at how a person's first language and their level of English proficiency might affect this relationship. By using detailed measures of musical ability, language proficiency, and differences between languages, this research aims to better understand how these factors interact when people perceive prosody in a second or foreign language.

We appreciate your interest in participating in this online task. You have been invited to participate as you are an adult English as a second language (ESL) or foreign language (EFL) speaker and have no language-related disorders. Please read through this information before agreeing to participate (if you wish to) by ticking the 'yes' box below.

You may ask any questions before deciding to take part by contacting the researcher (details below).

The Principal Researcher is Xinyi Pei, who is attached to the Department of Education at the University of Oxford. This research is being completed under the supervision of Elizabeth

Wonnacott.

You will be asked to complete the following tasks online: (1) a series of tasks evaluating your musical abilities across melody, tuning, rhythm, and tempo; (2) prosody perception tasks, where you will listen to audio recordings of sentences and complete text-matching tasks; (3) a quick lexical decision task to assess your English proficiency; and (4) a short background questionnaire about your age, gender, educational background, language background, and musical experience.

The tasks are expected to take about 50 minutes. No background knowledge is required. The data collected will be used to analyse how musical ability, first language and English proficiency influence prosody perception in English. The results will be primarily used in a master's dissertation and may be used in academic publications, conference presentations, and reports. Only the research team will have access to the raw data.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by pressing the 'Exit' button/ closing the browser. However, we are only able to reimburse participants who complete all research activities.

In the background questionnaire, we have included a 'Prefer not to say' option for each set of questions should you prefer not to answer a particular question.

How will my data be used?

We won't collect data that could directly identify you. We will collect your age (not DOB), gender (if you chose to share it) and some broad details about your educational background, the languages you speak, and your musical experience. We won't collect your IP address.

We will take all reasonable measures to ensure that data remain confidential.

The responses you provide will be stored in a password-protected electronic file on University of Oxford secure servers and may be used in academic publications, conference presentations, or reports for external organisations.

Who will have access to my data[1]?

The University of Oxford is the data controller with respect to your personal data and, as such, will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest. Further information about your rights with respect to your personal data is available from https://compliance.admin.ox.ac.uk/individual-rights.

The data you provide will be shared with the researcher and supervisor.

We would like to use the data in future studies, and to share data with other researchers (e.g. in online databases such as the open science framework (OSF). Data will have identifying information removed before it is shared with other researchers or results are made public.

The results will be written up for an MSc degree and possibly also for publication in academic journals. It may also be presented at conferences.

Who has reviewed this research?

This research has been reviewed by, and received ethics clearance through, a subcommittee of the University of Oxford Central University Research Ethics Committee [EDUC_1451961].

Who do I contact if I have a concern or I wish to complain?

If you have a concern about any aspect of this research, please contact Xinyi Pei at xinyi.pei@sjc.ox.ac.uk or Elizabeth Wonnacott at elizabeth.wonnacott@education.ox.ac.uk, and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the University of Oxford Research Governance, Ethics & Assurance (RGEA) team at rgea.complaints@admin.ox.ac.uk or on 01865 616480.

[1] Guidance is available within CUREC's Best Practice Guidance (09) on Data collection, protection and management

Please note that you may only participate in this survey if you are 18 years of age or over.

☐I certify that I am 18 years of age or over

If you have read the information above and agree to participate with the understanding that the data (including any personal data) you submit will be processed accordingly, please tick the box below to start.

☐Yes, I agree to take part

Appendix C - Background Information Questionnaire

1. What’s your gender? (Single choice)

   A. Female

   B. Male

   C. Non-binary

   D. Prefer not to say

2. What’s your age? (Open text)

3. What’s your nationality? (Open text)

4. What is your native language (mother tongue)? (Open text)

5. Apart from your native language, do you speak other languages? (Single choice)

   A. Yes

   B. No

6. Apart from your native language, what language(s) do you speak BEST? (Open text)

7. Apart from your native language, what language(s) do you speak and how proficient are you in these languages? (Open text, e.g., ‘language 1: [proficiency level], language 2: [proficiency level]. If none, write ‘NA’”)

   l Language 1:

   l Language 2 (if none, write “NA”):

   l Other (if none, write “NA”):

8. What is the level of your English? (Single choice)

   A. Native-like

B. Advanced/professional

C. Intermediate or fluent

D. Beginner

E. Prefer not to say

F. Other (please specify):

9. How frequently do you use English in your daily life? (Single choice)

A. Rarely (0-10%)

B. Occasionally (10-30%)

C. Sometimes (30-50%)

D. Often (50-70%)

E. Very often (70-90%)

F. Almost always (>90%)

10. At what age did you start learning English? (Open text, numerical)

11. Have you taken any standardized language proficiency tests (e.g., IELTS, TOEFL)? (Single choice with conditional open text)

A. No

B. Yes (please specify the test and your score):

12. Have you lived in an English-speaking country or region (e.g., UK, USA, Canada, Australia, etc.) for more than one month? (Single choice with conditional open text)

A. Yes

B. No

l If so, at what age did you live there, and for how long? (Open text)

13. Which of the following is your highest level of educational achievement? (Single choice)

A. No diploma (no formal education, primary, or lower secondary)

B. Upper secondary education (e.g., high school diploma, A-levels, IB diploma)

C. Post-secondary non-tertiary education (e.g., vocational/technical education)

D. Bachelor's degree or equivalent

E. Master's degree or equivalent

F. Doctoral degree (PhD) or equivalent

G. Other (please specify):

14. Would you consider yourself as a(n)... (Single choice)

A. Non-musician

B. Music loving Non-musician

C. Amateur musician

D. Semi-Professional musician

E. Professional musician

15. How often do you listen to music? (Single choice)

A. Never

B. Less than once a week

C. 1-2 days/week

D. 3-4 days/week

E. 5-6 days/week

F. Everyday

16. Do you play musical instrument(s) or sing? (Single choice with conditional open text)

A. Yes

B. No

l If yes, which instrument(s) do you play (including singing)? Please state how many years you have been playing the instrument(s)/have been singing. If you play more than three instruments, please state the ones that you play the longest. (Open text)

17. On average, how many hours do you play/practice per week? (Slider choice)

- Range: “0 hours/I no longer play” to “50 hours or more” (Participants select a value along a continuous scale)

18. Do you have any musical non-academic qualification(s) or music award(s)? (Single choice with conditional open text)

    A. No

    B. Yes (please specify):

19. Do you have a degree in applied music subjects (e.g., instrumental or vocal subjects, music education, composition, conducting) or music theory subjects? (Single choice with conditional options)

    A. Yes

    B. No

    - If yes, which kind of degree(s)? (Single choice)

        A. Bachelor

        B. Master / Diploma

        C. PhD / Doctorate

        D. Other (please specify):

(End of Questionnaire)

Appendix D - Power Analysis

The number of participants was determined based on the power the study hoped to achieve. According to Cohen (1988), power is the probability of correctly rejecting a false null hypothesis, balancing the risk of type I error (α) against type II error (β) (Bababekov et al., 2019). Setting α at 0.05 and β at 0.2 results in a conventional threshold for power of 0.8. However, Cohen’s (1988) generic effect size guidelines (small d=0.2, medium d=0.5, large d=0.8) have been shown to be less applicable in L2 research. Based on 346 primary studies and 91 meta-analyses, Plonsky and Oswald (2014) found that the distribution of observed effect sizes in L2 research differs substantially from these general benchmarks, thus proposed an updated field-specific threshold (small d=0.4, medium d=0.7, and large d=1.0) for between-group comparisons, and within-group contrasts and correlation coefficients (r=0.25 small, r=0.4 medium, r=0.6 large). Their field-specific benchmarks suggest that effect sizes in L2 research tend to be larger than those typically used in other fields, and medium effect sizes in L2 studies typically fall around d=0.7 for between-group designs and r=0.4 for correlational analyses. For multiple regression and correlational models, where the effect size metric is the squared semipartial correlation coefficient (f^2), a medium f^2 of 0.15 is consistent with r values around 0.3–0.4, as recommended by Plonsky and Oswald for correlations. For interaction models, a small effect size is required to assess interaction effects, which requires a f^2 of 0.06. This study thus adopts these suggested benchmarks by Plonsky and Oswald, and used the G*Power 3.1 programme (Faul et al., 2007) for power analysis to decide on the sample size.

For RQ1, which examines the extent to which music aptitude, L1, and language proficiency predict English prosody perception, a bivariate correlation analysis was planned. G*Power’s “Correlation: Bivariate normal model” was used to calculate the required sample size to detect an effect size of r=0.3 (medium by Plonsky & Oswald’s L2-specific thresholds) with 80% power, resulting in an estimated 84 participants.

Figure D.1

*G*Power analysis results for RQ1.*

For RQ2, which uses multiple regression to assess whether music aptitude predicts prosody perception while controlling for L1 distance and English proficiency, G*Power's "Linear multiple regression: Fixed model, R^2 deviation from zero" was used. With three predictors and a medium effect size (f^2=0.15, reflecting r=0.3–0.4), the required sample size was estimated at 77 participants.

Figure D.2

*G*Power analysis results for RQ2*

For RQ3, which explore how L1 distance and English proficiency moderate the relationship between music aptitude and prosody perception, linear mixed-effects models (LMMs) were used. The model included three main predictors—music aptitude, L1 distance, and English proficiency and two interaction terms (music aptitude × L1 distance and music aptitude × English proficiency), resulting in a total of five predictors. Power analyses were conducted separately for each interaction term using G*Power's "Linear multiple regression: Fixed model, R^2 increase" function, specifying one tested predictor (the interaction term) and four total predictors. For a small effect size ($f^2 = 0.06$) and power of 0.8, the required sample size for detecting each interaction effect was estimated at approximately 133 participants.

Figure D.3

*G*Power analysis results for RQ3*

For RQ4, which examines which aspects of prosody perception are most influenced by music aptitude and how these effects vary by L1 distance and English proficiency, linear mixed-effects models (LMMs) were used. The independent variables included the separate subtest scores of music aptitude scores (Mini-PROMS), L1 distance, and LexTALE scores. The dependent variables were two dimensions prosody perception accuracy (pitch and time), measured by the total scores of pitch-only condition and timing-only condition.

Each set of models included the main effects of a subtest of music aptitude, L1 distance, and English proficiency, as well as interaction terms between music aptitude and the two moderators (L1 distance and English proficiency), resulting in four iterations of the same model per prosody dimension (pitch and timing), and a total of eight models.

Power analyses for detecting interaction effects in these models were conducted separately for each interaction term using G*Power’s ‘Linear multiple regression: Fixed model, R^2 increase” function. Each analysis specified one tested predictor (the interaction term) and four total predictors (three main effects plus the interaction). Assuming a medium effect size ($f^2 = 0.06$) and power of 0.8, the required sample size was estimated at approximately 133 participants per interaction test.

Figure D.4

*G*Power analysis results for RQ4*

In summary, the power analyses indicate that a total of 133 participants would be required to ensure adequate power across all research questions. However, due to reality constraints, a total of 100 participants were involved in the study, which, though adequate for RQ1 and RQ2, is slight underpower to fully address RQ3 and RQ4.

Appendix E - LexTALE Words List and Answer

1	mensible	0	21	slain	1	41	nourishment	1
2	scornful	1	22	recipient	1	42	abergry	0
3	stoutly	1	23	exprate	0	43	proom	0
4	ablaze	1	24	eloquence	1	44	turmoil	1
5	kermshaw	0	25	cleanliness	1	45	carbohydrate	1
6	moonlit	1	26	dispatch	1	46	scholar	1
7	lofty	1	27	rebondicate	0	47	turtle	1
8	hurricane	1	28	ingenious	1	48	fellick	0
9	flaw	1	29	bewitch	1	49	destription	0
10	alberation	0	30	skave	0	50	cylinder	1
11	unkempt	1	31	plaintively	1	51	censorship	1
12	breeding	1	32	kilp	0	52	celestial	1
13	festivity	1	33	interfate	0	53	rascal	1
14	screech	1	34	hasty	1	54	purrage	0
15	savoury	1	35	lengthy	1	55	pulsh	0
16	plaudate	0	36	fray	1	56	muddy	1
17	shin	1	37	crumper	0	57	quirty	0
18	fluid	1	38	upkeep	1	58	pudour	0
19	spaunch	0	39	majestic	1	59	listless	1
20	allied	1	40	magrity	0	60	wrought	1

Note: 0 = non-word, 1 = real word in English.

Appendix F - Prosody Perception (MBOPP) Stimuli

1. Prosodic Focus Subtest Stimuli

	Start	Focus 1	Focus 2	Middle	Ending 1	Ending 2
Example	Mary likes to	read	books	but she doesn't like to	WRITE books	read MAGAZINES
Practice 1	Alice sometimes	pets	dogs	but she won't	WASH dogs	pet cats
Practice 2	Dave likes to	study	music	but he doesn't like to	PLAY music	study HISTORY
Test 1	Sally has a	Windows	computer	but she really wants	an APPLE computer	a Windows TABLET
Test 2	George asked for a	white	Americano	but the barista gave him a	BLACK Americano	white filter COFFEE
Test 3	Fiona was eating	strawberry	yoghurt	but she really wanted some	BLUEBERRY yoghurt	strawberry ICECREAM
Test 4	Tom likes	barbecue	chicken	but not as much as	ROAST chicken	barbecue PORK
Test 5	Sophie likes to	paint	landscapes	but she doesn't like to	DRAW landscapes	paint PORTRAITS
Test 6	John can't	run	a marathon	but he could	WALK a marathon	run a MILE
Test 7	Matt is good at	flying	planes	but he isn't good at	LANDING planes	flying HELICOPTERS
Test 8	Pippa found a	jam	jar	but she couldn't find a	JELLY jar	jam KNIFE
Test 9	Sam has a	fish	knife	but he doesn't have a	BUTTER knife	fish FORK
Test 10	Rachel likes	French	food	but she doesn't like	ITALIAN food	French WINE
Test 11	The woman likes	white	pearls	but not	BLACK pearls	white DIAMONDS
Test 12	Ken won't buy	Sainsbury's	pizza	but he will buy	TESCO'S pizza	Sainsbury's CHICKEN
Test 13	Sarah has a	Barclay's	card	but she doesn't have a	LLOYDS card	BARCLAY'S MORTGAGE
Test 14	Neil won't support	Oxford's	fencing	but he will support	CAMBRIDGE'S fencing	Oxford's ROWING team
Test 15	Carolyn likes	Scottish	pubs	but she doesn't like	ENGLISH pubs	Scottish RESTAURANTS
Test 16	Micah been to	Regent's	park	but he hasn't been to	HYDE PARK	Regent's STREET
Test 17	Rosalyn likes to	drink	beer	but she doesn't like to	BREW beer	drink LIQUOR
Test 18	Veronica has visited	America	for holiday	but she hasn't visited	CANADA for holiday	AMERICA FOR WORK
Test 19	Tim has an	electric	piano	but he really wants an	ACOUSTIC piano	electric GUITAR
Test 20	Ben has ridden a	UK	train	but he has never ridden a	AMERICAN train	UK BUS
Test 21	Nancy has a	small	flat	but she would really like a	LARGE flat	small HOUSE
Test 22	Paul's house has a	brown	sofa	but it doesn't have a	BLACK sofa	brown CHAIR

2. Phrase Boundary Subtest Stimuli

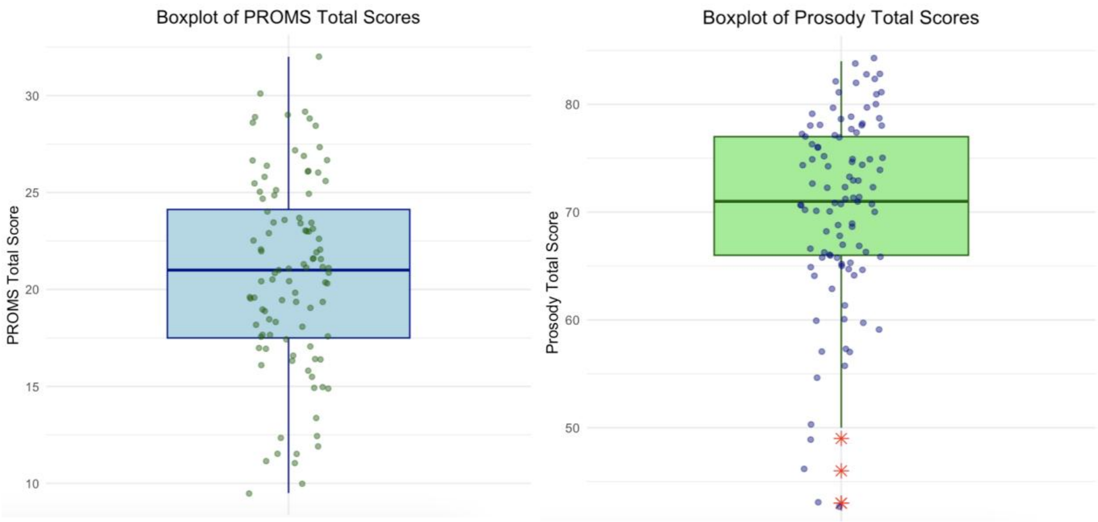
	Closure	Start	Ending
Example	Early	After Jane dusts, the dining table	is clean
Example	Late	After Jane dusts the dining table,	it's clean
Practice 1	Early	After John runs, the race	is over
Practice 1	Late	After John runs the race,	it's over
Practice 2	Early	Because Mike phoned, his mother	was relieved
Practice 2	Late	Because Mike phoned his mother,	she was relieved
Test 1	Early	Because Sarah answered, the teacher	was proud
Test 1	Late	Because Sarah answered the teacher,	she was proud
Test 2	Early	Because Tara cleaned the house,	was spotless
Test 2	Late	Because Tara cleaned the house,	it was spotless
Test 3	Early	Because George forgot, the party	had started
Test 3	Late	Because George forgot the party,	he was sad
Test 4	Early	Because Mike paid, the bill	was smaller
Test 4	Late	Because Mike paid the bill,	it was smaller
Test 5	Early	If Charles is baby-sitting, the children	are happy
Test 5	Late	If Charles is baby-sitting, the children,	they're happy
Test 6	Early	If George is programming, the computer	is busy
Test 6	Late	If George is programming the computer,	it's busy
Test 7	Early	If Ian doesn't notice, Beth	is fine
Test 7	Late	If Ian doesn't notice, Beth,	it's fine
Test 8	Early	If Joe starts, the meeting	will be long
Test 8	Late	If Joe starts the meeting,	it'll be long
Test 9	Early	If Laura is folding, the towels	will be neat
Test 9	Late	If Laura is folding the towels,	they'll be neat

	Closure	Start	Ending
Test 10	Early	When the baby finishes, the bottle	will be empty
Test 10	Late	When the baby finishes the bottle,	will be empty
Test 11	Early	If Barbara gives up, the ship	will be plundered
Test 11	Late	If Barbara gives up the ship,	it'll be plundered
Test 12	Early	If the Scissor Sisters open the show,	will be great
Test 12	Late	If the Scissor Sisters open the show,	it'll be great
Test 13	Early	If the maid packs, the suitcase	will be tidy
Test 13	Late	If the maid packs the suitcase,	it'll be tidy
Test 14	Early	If Tom wins, the contest	is over
Test 14	Late	If Tom wins the contest,	it's over
Test 15	Early	If the doctor calls, your sister	will answer
Test 15	Late	If the doctor calls, your sister,	she'll answer
Test 16	Early	If Jack cleans, the kitchen	will be filthy
Test 16	Late	If Jack cleans the kitchen,	it'll be filthy
Test 17	Early	If dad digs, the hole	will be deep
Test 17	Late	If dad digs the hole,	it'll be deep
Test 18	Early	When a man cheats, his friends	get angry
Test 18	Late	When a man cheats, his friends,	they're angry
Test 19	Early	When Gaga sings, the song	is a hit
Test 19	Late	When Gaga sings the song,	it's a hit
Test 20	Early	When Roger leaves, the house	is dark
Test 20	Late	When Roger leaves the house,	it's dark

Appendix G - Outliers

Figure G.1

Boxplots for Outliers in PROMS Total Scores and Prosody Total Scores



Outlier detection was performed for continuous variables using the inter-quartile range (IQR) method (data points falling outside Q1−1.5× IQR or Q3+1.5× IQR).

In both boxplots shown above, the box consists of three horizontal lines, which represent Q1 (the first quartile), Q2 (median), and Q3 (the third quartile). The range represents the inter-quartile range. Each dot represents one data point. The dots that scatter outside the Q3 are considered outliers. By calculating the outlier boundaries in R, four outliers were found in the Prosody Total scores. However, as these scores represent genuine individual differences in prosody perception, rather than data entry errors or measurement flaws, all four outliers were retained in the study.

Appendix H - Normality Check for Data Analysis

2.1 Normality Check for Key Variables (Total Scores)

2.1.1 Shapiro-Wilk Test

The Shapiro-Wilk test results, summarised in Table 2.1, provided initial insights into the normality of the data distributions.

Table H.1

Shapiro-Wilk Results

Variable	W Value	p-value	Interpretation
PROMS Total	0.98705	0.4409	p > 0.05, data follows normal distribution
Prosody Total	0.93663	0.0001208	p < 0.05, data does not follow normal distribution
LexTALE	0.97305	0.0380	p < 0.05, data does not follow normal distribution
L1 Distance	0.79762	2.093e-10	p << 0.05, data significantly deviates from normal distribution

As shown in Table H.1, ***PROMS Total*** (W = 0.98705, p = 0.4409) demonstrated a normal distribution (p > 0.05). Conversely, ***Prosody Total*** (W = 0.93663, p = 0.0001208), ***LexTALE*** (W = 0.97305, p = 0.0380), and ***L1 Distance*** (W = 0.79762, p = 2.093e-10) all showed statistically significant deviations from a normal distribution (p < 0.05).

However, it is crucial to note that, given the sample size of 100 in this study, which is considered a large sample size, the Shapiro-Wilk test can be overly sensitive. They may detect minor deviations from a normal distribution, labeling the data as non-normal, even when these deviations are too small to have a meaningful impact on the results of subsequent statistical analyses (Field et al., 2012). Therefore, a definitive interpretation of true normality requires a combined assessment with visual inspections (histograms and Q-Q plots) and an examination of skewness and kurtosis values, which provide a more reliable understanding of the data's distribution.

2.1.2 Skewness and Kurtosis

Further supplementing the Shapiro-Wilk findings, the analysis of skewness and kurtosis provided additional quantitative measures of the shape of the data distributions. Table 2.2 presents these descriptive statistics for the total scores of the key variables.

Table H.2

Skew and Kurtosis Results

Variable	Skew	Kurtosis	z_skew	z_kurtosis
PROMS Total	-0.18	-0.44	-0.73	-0.90
Prosody Total	-0.94	0.81	-3.83	1.66
LexTALE	-0.36	-0.39	-1.48	-0.80
L1 Distance	-0.049	-1.62	-0.2	-3.31

As observed in Table H.2, *PROMS Total* displayed relatively low absolute values for skewness (-0.18) and kurtosis (-0.44), with corresponding z-scores of -0.73 and -0.90. These z-scores are well below the ±1.96 threshold for significance at $p < .05$, meaning that they are well within the commonly accepted range for approximate normality (Field et al., 2012). further supporting the notion that *PROMS Total* is acceptably normally distributed, as initially suggested by the Shapiro-Wilk test.

In contrast, *Prosody Total* showed a moderate negative skewness (-0.94) and positive kurtosis (0.81), with z-scores of -3.83 and 1.66. The z-skew value of -3.83, being greater than ±2.58 (significant at $p < .01$), clearly indicates a statistically significant deviation from symmetry.

For *LexTALE*, despite the Shapiro-Wilk test showing a statistically significant result ($p = 0.0380$), a closer examination of its skewness and kurtosis z-scores (-1.48 and -0.80 respectively) reveals that both values fall within the ±1.96 threshold for significance in a sample of this size, suggesting that *LexTALE* can be considered to conform to an acceptable normal distribution.

Finally, *L1 Distance* presented a very small negative skewness (-0.049) but a notable negative kurtosis (-1.62), with z-scores of -0.2 and -3.31. The large negative kurtosis z-score of -3.31, which is above ±3.29 (the threshold for significance at $p < .001$; Field et al., 2012), indicates a significantly flatter distribution than a normal one, implying a wider dispersion of scores around the mean and a lack of clear central tendency.

2.1.3 Visual Inspections

Visual inspections of histograms and Q-Q plots provide a graphical representation of the distribution characteristics, offering a qualitative complement to the statistical tests. Figure 2.1 and Figure 2.2 display the histograms and Q-Q plots for the key variables—music aptitude scores (presented as *PROMS Total*), prosody perception scores (presented as *Prosody Total*), English proficiency scores (presented as *LexTALE*), and L1 language distance (presented as *L1 Distance*). Since a quick visual inspection shows that data of L1 language distance is clearly not normally distributed, it is excluded from the Q-Q plot, while the other three variables are presented in the Q-Q plot to have a detailed inspection of its distribution.

Figure H.1

Histograms of Variable Distributions

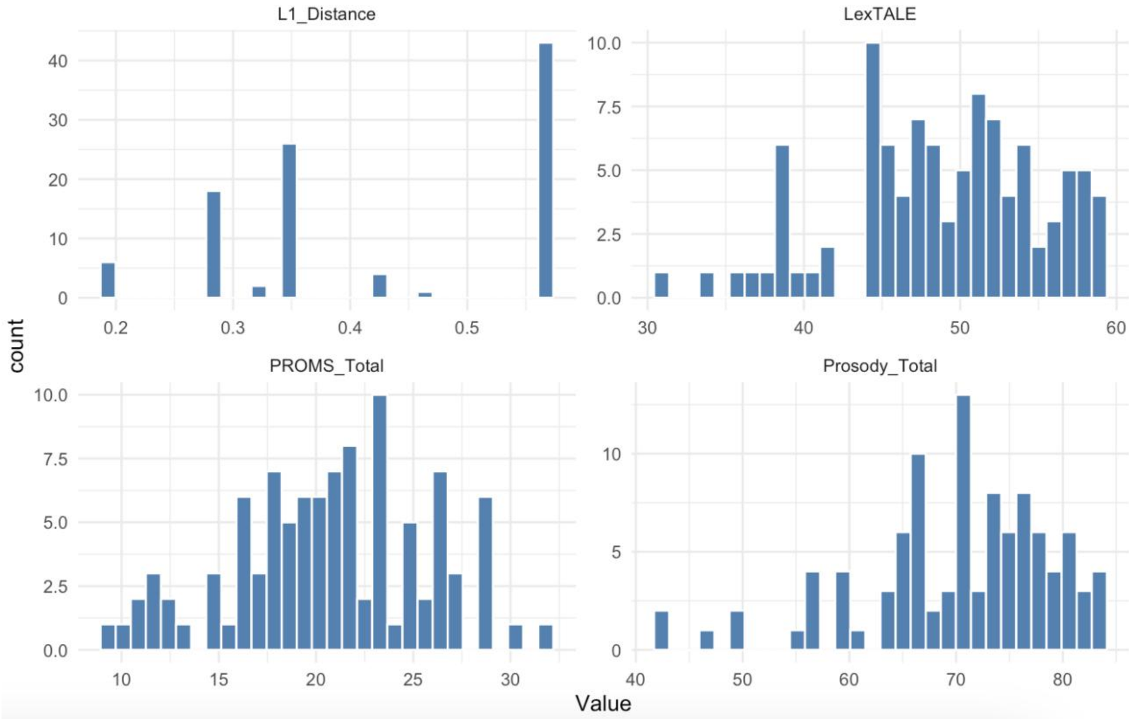
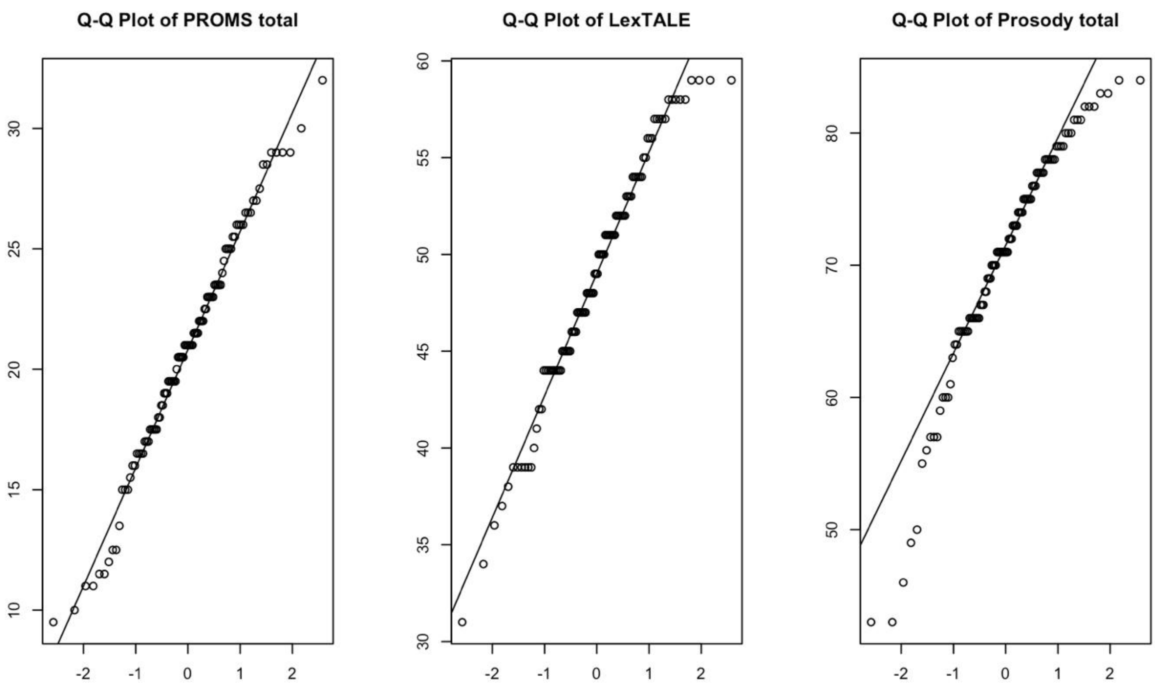


Figure H.2

Q-Q Plot of Variable Distributions



The histogram for *PROMS Total* (Figure H.1) visually displays a roughly symmetrical, bell-shaped distribution, with most scores clustering around the mean. This visual impression is strongly corroborated by its Q-Q plot (Figure H.2), where the data points align very closely along the theoretical normal distribution line. This close alignment confirms that *PROMS Total* data adheres well to a normal distribution, consistent with the non-significant Shapiro-Wilk test and the low skewness and kurtosis z-scores.

The *LexTALE* histogram (Figure H.1) appears generally symmetrical, though with some minor variations in bar height that do not indicate a severe departure from normality. Furthermore, its Q-Q plot (Figure H.2) shows that the data points generally follow the straight line, with only slight deviations, particularly at the lower and upper tails. Considering these visual insights alongside the non-significant skewness and kurtosis z-scores (despite the sensitive Shapiro-Wilk test result), the visual evidence supports the conclusion that *LexTALE* conforms to an acceptable normal distribution for the purpose of further analysis.

The histogram for *Prosody Total* (Figure H.1) shows a leftward skew, with a higher frequency of scores towards the upper end and a tail extending towards lower values. This visual indication of negative skewness is strongly supported by its Q-Q plot (Figure H.2), where the data points deviate significantly from the straight line at both tails. These visual findings consistently confirm the non-normal distribution identified by the Shapiro-Wilk test and the significant negative skewness z-score.

The histogram for *L1 Distance* (Figure H.1) displays a distinctly non-normal distribution, appearing rather flat and showing several distinct peaks. This visual evidence of a lack of a clear central tendency reinforces the significant negative kurtosis identified by the statistical tests, indicating a significant deviation from a normal distribution.

2.2 Normality Check for Residuals

Before interpreting the regression coefficients for RQ2, RQ3 and RQ4, the assumptions of the multiple linear regression model were examined. Visual inspection of the residuals through a histogram and a Q-Q plot (Figure H.3) indicated a deviation from normality. Specifically, the Q-Q plot showed the residuals not strictly following the diagonal line, particularly at the tails. Quantitative tests further supported this observation. The Shapiro-Wilk normality test on the residuals yielded a significant result (W=0.95599, p=0.002103), leading to the rejection of the null hypothesis that the residuals are normally distributed. Furthermore, the skewness of the residuals was −3.437729 (Z-score for skewness = −3.437729), indicating a significant negative (left) skew. The kurtosis of the residuals was 8.576266 (Z-score for kurtosis = 8.576266), suggesting a significantly leptokurtic (peaked with heavy tails) distribution. These results collectively confirmed that the residuals did not conform to the assumption of normality, thus underscoring the necessity of using bootstrapping, a powerful, non-parametric resampling technique that does not require assumptions about data distribution like normality (Field et al., 2012), as a supplementary method for robust inference.

Figure H.3

Histogram and Q-Q Plot for Normality Check of the Residuals