Palacký University Olomouc

Faculty of Arts

Department of English and American Studies

The Impact of Musical Aptitude on Czech Learners’ Perception of L2 English Lexical Stress

(Master thesis)

2024

Bc. Helena Rychetská

The Impact of Musical Aptitude on Czech Learners' Perception of L2 English Lexical Stress

(Diplomová práce)

Autor: Bc. Helena Rychetská

Studijní obor: Anglická filologie + Muzikologie

Vedoucí práce: Mgr. Václav Jonáš Podlipský, Ph.D.

Počet znaků: 138 238 (bez appendixů)

Počet stran (podle znaků /1800): 78

Počet stran (podle čísel): 73

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a uvedla úplný seznam citované a použité literatury.

V Olomouci dne 27. 6. 2024 **Bc. Helena Rychetská**

Abstract: The primary objective of this thesis is to investigate the correlation between musical aptitude and the perception of English lexical stress among Czech speakers of L2 English. The study utilizes the Profile of Music Perception Skills - PROMS-S (Law and Zentner, 2012) to measure musical aptitude and assesses English stress perception through an experiment based on Tlolková (2018). Participants, consisting of 54 university students of Musicology or English Philology, are evaluated on their ability to perceive English stress, their musical aptitude across various domains (melody, rhythm, rhythm- melody, tuning, beat, timbre, tempo, and pitch), and their English proficiency as indexed by the LexTALE vocabulary test (Lemhöfer & Broersma, 2012).

The research involves three assessments: (1) an AXB discrimination task testing English stress perception using pseudoword stimuli with variations in acoustic cues (pitch, intensity, duration, vowel formants, and spectral tilt), (2) the PROMS-S battery measuring musical aptitude in multiple perceptual domains, and (3) the LexTALE test determining the participants' English vocabulary size.

Results from a mixed-effects logistic regression model reveal that musical perception skills significantly predict the accuracy of L2 English stress perception. Specifically, temporal and dynamic musical skills, such as rhythm, tempo, and beat, enhance stress perception, while tonal and spectral skills do not have a reliable impact. The interaction between musical aptitude and vocabulary size suggests that the highest accuracy in stress perception is achieved by participants with high scores in both areas.

In conclusion, the findings indicate that even advanced Czech learners of English experience challenges with "stress deafness." However, strong temporal and dynamic musical abilities substantially improve their perception of L2 lexical stress, highlighting the importance of specific musical components in language acquisition and their relationship with English language proficiency.

Keywords: Stress deafness, perception of English lexical stress, musical aptitude, language proficiency

Anotace: Cílem této diplomové práce je prozkoumat vztah mezi hudebním nadáním a vnímáním přízvuku v angličtině u českých rodilých mluvčích, jejichž druhým jazykem je angličtina. Experiment využívá test hudebních schopností Profile of Music Perception Skills - PROMS-S (Law a Zentner, 2012) k měření hudebního nadání a hodnotí vnímání anglického přízvuku prostřednictvím experimentu navrženého Tlolkovou (2018). Studie zahrnuje data od 54 univerzitních studentů muzikologie nebo anglické filologie, u kterých je hodnocena jejich schopnost vnímat anglický přízvuk, hudební kompetence v rozličných oblastech (melodie, rytmus, rytmus-melodie, ladění, beat, barva zvuku, tempo a výška tónu) a také úroveň angličtiny, jež se měří testem slovní zásoby LexTALE (Lemhöfer a Broersma, 2012).

Experiment zahrnuje tři testy: (1) AXB diskriminační test, ověřující vnímaní anglického přízvuku na neexistujících anglických slovech, jejichž stimuly se liší v akustických parametrech (výška tónu, intenzita, trvání, formanty samohlásek a spektrální náklon), (2) PROMS-S baterie měřící hudební schopnosti v několika percepčních oblastech a (3) test LexTALE určující velikost slovní zásoby účastníků.

Výsledky smíšeného logistického regresního modelu ukazují, že hudební percepční schopnosti významně predikují přesnost vnímání přízvuku v angličtině jako druhém jazyce. Časové a dynamické hudební složky, měřené v rámci PROMS-S, jako je rytmus, tempo a beat, zlepšují vnímání přízvuku, zatímco tonální a spektrální složky na něj nemají spolehlivý vliv. Interakce mezi hudebními schopnostmi a velikostí slovní zásoby naznačuje, že nejlépe vnímají přízvuk účastníci, kteří dosáhli vysokých výsledků v obou oblastech.

Z provedeného experimentu vyplývá, že i pokročilejší čeští studenti angličtiny mají problémy s rozeznáváním přízvuku v anglickém jazyce, avšak lepší hudební nadání v časových a dynamických doménách výrazně zlepšuje schopnost rozlišování přízvuku v angličtině. Z toho vyplývá, že hudební nadání, ve spojení s úrovní znalosti angličtiny, hraje při osvojování cizího jazyka významnou roli.

Klíčová slova: Neschopnost vnímat přízvuk, percepce přízvuku v angličtině, hudební nadání, jazyková způsobilost

Acknowledgements

I would like to express my gratitude to my supervisor, Mgr. Václav Jonáš Podlipský, Ph.D., for his patience, the time he dedicated to guiding me through my thesis and for his work on Praat scripts and modelling in R. Also, many thanks to all who participated in the experiment.

TABLE OF CONTENTS

1 Introduction

In the realm where music and language converge, this thesis delves into the intricate relationship between musical aptitude and second language (L2) pronunciation, with a specific focus on English stress perception. Recognizing the shared fundamental characteristics of music and language, both of which are processed by the auditory system, this study sheds light on their intertwined nature. Auditory properties such as pitch, loudness, quality (or timbre), and timing are important in shaping human perception in speech and music alike. When we place a sound on a scale as lower or higher, distinguishing different pitches, we perceive variations in frequency, reflecting the rates of vibration in air pressure. Similarly, we differentiate sounds based on their loudness, envisioning striking an instrument gently or with force, which corresponds to the movement of the eardrum in human perception. Quality of sounds (or timbre, as it is termed in music) refers to distinct sound characteristics, while timing pertains to the temporal events of sounds (Ladefoged 1996, 14-16). These auditory properties are particularly crucial in deciphering language suprasegmentals like stress. In the forthcoming chapter on music and language, the intersection of these shared auditory properties with linguistic suprasegmentals will be explored, focusing on stress cues in L2 English perception and identifying pertinent musical components.

This study aims to contribute to the existing body of knowledge by investigating the impact of musical aptitude on stress perception by Czech learners of English. While researchers (e.g. Culp 2017; Gralińska-Brawata and Rybińska 2017; Milovanov et al. 2010; Talamini et al. 2018) acknowledge the superior abilities of musicians to learn L2 sound patterns and emphasize the influence of musical skills on L2 learning, the role of musical aptitude among Czech learners remains relatively unexplored. This thesis addresses this gap through an experimental approach.

The primary goals of this thesis involve assessing the correlation between musical aptitude, measured using the standardized musical aptitude perception test Profile of Music Perception Skills - PROMS-S (Law and Zentner 2012), and English stress perception in an experiment devised by Tlolková (2018) with Czech speakers of L2 English, recruited amongst Musicology and English Philology students. Furthermore, this thesis investigates connections between musical ability in specific domains (melody, rhythm, rhythm-melody, tuning, beat, timbre, tempo, and pitch) and English stress perception, while also considering the participants' English proficiency as indexed by LexTALE, a lexical decision task providing an estimate of vocabulary size (Lemhöfer & Broersma 2012). This investigation aims to identify the specific musical components crucial to L2 English stress perception and to comprehend their potential correlation with English language proficiency.

As highlighted by Zybert and Stepień (2009, 100), music and language share an expressive nature, with speech expressiveness relying on suprasegmentals, i.e. intonation (patterns of changing the pitch of the voice) and rhythm (patterns of recurring stresses with different timing). This thesis centres on stress, particularly because Czech learners often face challenges both in producing English stress and perceiving it and could be described as "stress deaf," given the contrastive nature of stress in English, which varies in placement, unlike the non-lexical and phonetically weak stress fixed on the first syllable of a word in Czech (Skarnitzl and Rumlová 2019). Additionally, the unstressed syllable in English is typically reduced, a feature absent in the Czech language (Skarnitzl and Rumlová 2019, 113). The objective of the present thesis is to test whether Czech learners of L2 English with higher musical aptitude exhibit superior stress perception compared to those with lower musical aptitude.

Central to this work is the concept of musical aptitude, defined by Talamini et al. (2018, 2) as "[…] a talent for perceiving and discriminating musical sounds, such as melodies, chords, rhythms, and so on." This notion, intricately linked to genetic predispositions, is posited by Culp (2017, 330) to potentially influence how readily students grasp and comprehend musical information. According to Gordon (2007, 44), "musical aptitude is a measure of one's potential or capacity to learn music," suggesting that every individual possesses some degree of musical ability. Gordon's (2007) analysis reveals that approximately two-thirds of individuals possess average aptitude in music, with the remaining population distributed above or below this norm, and only a small fraction exhibiting extremely high or low levels of aptitude. He draws a critical distinction between music achievement, which involves intellectual processes in the brain, and musical aptitude, which is rooted in genetic predispositions and encompasses the entire body. Musical aptitude is shaped by innate potential and environmental influences, particularly during the developmental musical aptitude stage, which extends until around the age of nine (Gordon 2007, 47). Consequently, an individual's capacity to excel in music tends to stabilize around the age of nine and persists into adulthood. Those with high musical aptitude may fail to fully realize their potential without adequate instruction or training, while individuals with lower aptitude may surpass those with average abilities if provided with appropriate guidance.[1] The standardized musical aptitude perception test PROMS-S (Law and Zentner 2012) evaluates the musical aptitude of participants at the time of testing. However, it is challenging to determine the extent to which their test scores are influenced by innate potential, environmental factors, or musical training/instruction, as participants were not tested before age nine and a questionnaire about their musical training may not fully capture the reasons for their musical aptitude scores. Carroll (1993, 376) suggests that musical aptitude tests rely on basic discriminations among tonal materials often lacking musical contexts, aiming to minimize the impact of musical training on test outcomes, and consequently making them predictive of success in such training. Nonetheless, developing tests with meaningful musical contexts without being unduly influenced by musical training poses a significant challenge (Carroll 1993, 376). There is a significant debate about whether differences in behaviour are attributable to inherent abilities or the amount and quality of practice (Christiner and Reiterer 2018, 9). Therefore, in this thesis, the term musical aptitude refers to the participants' musical ability at the time of testing, offering a snapshot of their innate abilities without accounting for their prior musical training or experience, and exploring its correlation with English stress perception. Nonetheless, individuals with innate musical talent are more likely to further develop their abilities successfully, potentially leading to an enhanced perception of L2 sounds over time (Carroll 1993).

The structure of this master's thesis is outlined as follows. The initial section explores the correlation between music and language, reviewing previous studies and highlighting their commonalities. Additionally, various musical aptitude tests used throughout the 20th and 21st centuries are examined. The focus then shifts to the perception of English stress, discussing the difference between production and perception, examining stress acoustic cues, and comparing stress patterns in English and Czech.

The second part of the thesis provides a detailed account of the experiment conducted with 54 Czech bachelor students from Palacký University in Olomouc, primarily comprising English Philology and Musicology students. The primary objective was to investigate the potential relationship between musical aptitude, English stress perception and English

[1] For further discussion on the impact of musical training on musical aptitude, refer to Nardo and Reiterer (2009), Pastuszek-Lipińska (2008) or Gordon (2007).

language proficiency, aiming to determine if high musical aptitude influences the English stress perception of Czech learners. The data were obtained from three tests: the English stress perception test from Tlolková's thesis (2018), additionally examining the acoustic cues to English stress in her stimuli, the music aptitude test PROMS-S (Law and Zentner 2012), and the lexical decision task (Lemhöfer & Broersma 2012). The results of statistical analyses are presented in chapter three followed by a general discussion and conclusion.

1.1 Music and Language

Music and language exhibit various parallels and distinctions. In terms of similarities, both are intrinsic to human experience and possess multiple forms of expression, including vocal, gestural, and written modes. Additionally, speech and song often develop concurrently, reflecting their shared reliance on auditory perception within the natural environment. Both possess the capacity for infinite creativity through the arrangement of musical contours or words, emphasizing the significance of rhythm and melody as fundamental components. Additionally, another common trait is their expressiveness, with speech relying on suprasegmental features such as intonation, rhythm, pitch, and stress. Moreover, they adhere to structured orders, whether through linguistic grammar or musical notes, fostering receptivity and productive abilities within cultural contexts (Fonseca-Mora et al. 2011, 102).

However, differences also abound. While language necessitates translation for cross-cultural understanding, the music defies such translation due to its cultural specificity. While individuals may be proficient in only one language, exposure to diverse musical styles is widespread. Additionally, while grammar lends meaning to language, musical rules lack inherent semantic content. The evolution of musical styles occurs more rapidly compared to language grammar, which evolves at a slower pace. Furthermore, unlike language, music lacks a direct manipulative function concerning real-world entities and relations. Cognitive demands differ between the two, with music posing fewer cognitive challenges than language. Temporal and rhythmic structures play vital roles in both domains, yet music's metric structure is specific to each piece, while language prosody exhibits greater variability. Finally, while the number and variety of tones remain consistent across cultures, the phonemic diversity of languages varies significantly (Fonseca-Mora et al. 2011, 102).

In this thesis, the focus primarily rests on the shared domain of auditory properties, given their intricate role in both music and language. Fonseca-Mora et al. (2011, 103) eloquently describe this concept, stating that "each language possesses distinct musical elements in the form of its prosody, encompassing intonation patterns and rhythmic structures." Similarly, Gralińska-Brawata and Rybińska (2017, 268) propose that both music and speech prosody exhibit resemblances, such as mutual elements like melody (intonation) and rhythm (stress and timing). This assessment aligns with Marques et al.'s (2007, 1454) that "[…] prosody relies on the same acoustic parameters (i.e. F0, duration, intensity, and spectral characteristics) as the melody […] of a musical phrase." Acoustically, music and speech share properties like frequency, duration, and intensity. The PROMS-S test used in the experiment assesses various receptive musical abilities in several areas, including tonal (such as melody and pitch), qualitative (like timbre and tuning), temporal (involving rhythm, rhythm-to-melody, accent, and tempo), and dynamic (covering loudness) domains. This alignment underscores the intricate relationship between musical and linguistic elements, particularly in the perception of sound sequences. Thus, through the PROMS-S test, insights into participants' abilities to discern and interpret auditory stimuli are gained, shedding light on the parallel processing mechanisms underlying both music and language perception.

An essential aspect requiring consideration is the correlation between the perception and production in both music and language, representing another commonality between them. This applies to both music and L2 acquisition, where individuals must learn or practice to reproduce certain phonemes in language or execute specific sounds when singing or playing an instrument in music. The assumption, that enhanced perceptual abilities lead to improved production, has been both supported and refuted (Jekiel and Malarski 2021, 1668). Moreover, recent findings suggest a bidirectional relationship where production can also shape perception. This suggests that language learners need consistent exposure to genuine spoken language patterns to enhance their perceptual learning abilities; otherwise, engaging in production activities could interfere with this progression (Jekiel and Malarski 2021, 1668). Without a strong foundation in perceiving and understanding the language, attempts to produce it may be less effective or even disruptive to the learning process. Individuals may exhibit divergent capacities in reception and production, as proficiency in one doesn't necessarily translate to proficiency in the other. A study by Zybert and Stepień (2009) identified a correlation between both domains, highlighting a stronger association between music perception and speech perception compared to music production and speech production. Additionally, their research suggests a positive correlation between the perception and production of speech, implying that the ability to recognize speech sounds also entails the capability to replicate them through imitation (Zybert and Stepień 2009, 109). Receptive pronunciation denotes the level of proficiency in processing language sounds (Rodríguez 2021, 75-76). Moreover, this thesis delves into phonological awareness, defined by Culp (2017, 329) as "a person's ability to analyse and manipulate the sounds of language, including units within words and/or entire words." From a basic standpoint, phonological awareness exhibits similarities with receptive musical aptitude, particularly in aspects like tone and timbre, as they both involve the processing and differentiation of sounds (Rodríguez 2021, 76). While existing studies primarily focus on the tonal domain in exploring the relationship between phonological awareness and musical aptitude, this thesis aims to broaden the scope. For instance, Culp's (2017, 340) study suggests that phonological awareness might have a stronger correlation with the tonal aspect rather than the rhythmic aspect. Additionally, Rodríguez (2021, 85) found that phonological awareness is connected to every musical subdomain, but most significantly with pitch and duration, given their essential role in English language phonetics. This expanded perspective offers a nuanced understanding of the interplay between phonological awareness and musical aptitude, shedding light on their complex relationship.

In exploring the intricate relationship between music and language, the subsequent chapters offer a comprehensive review of research in this area. The overview begins with an examination of linguistic studies, particularly those centred on stress perception, drawing insights from studies by Gralińska-Brawata and Rybińska (2017), Balčytytė-Kurtinienė (2015), and Garcia and Schwab (2023). Following this, the neuroscientific domain is delved into, where the neural underpinnings of the musical and linguistic interplay are unveiled, shedding light on the intricate connections between the two domains. Lastly, pedagogical approaches to music and language are scrutinized, exploring innovative methodologies and studies that illuminate how their shared elements can be effectively leveraged for language instruction.

1.1.1 Musical aptitude and L2 word stress

The intricate nature of stress placement in English words, with its significant impact on word meanings, prompts researchers to explore its correlation with musical abilities. Understanding the complexities of stress in English is crucial for language learners and educators alike, as it influences word comprehension and pronunciation accuracy. The studies exploring word stress include those by Gralińska-Brawata and Rybińska (2017), Balčytytė-Kurtinienė (2015), and Garcia and Schwab (2023).

Gralińska-Brawata and Rybińska's (2017) study, conducted with Polish learners, aimed to explore the relationship between their musical abilities and word stress production in English. Additionally, it investigated how this relationship influenced the teachability of word stress to advanced Polish students of English. Using auditory recordings, questionnaires, and a music test, the study analysed the data descriptively. Results indicated that students tended to overgeneralize word stress rules in English rather than applying the penultimate syllable rule from Polish. Moreover, a correlation between word stress production and participants' musical abilities was observed for most of the participants. This finding aligns with the traditional assumption that individuals with a music talent may have better pronunciation skills due to their heightened sensitivity to pitch variations and ability to imitate rhythmic patterns effectively (Gralińska-Brawata and Rybińska 2017, 265-266). Furthermore, Gralińska-Brawata and Rybińska's (2017) study involved a questionnaire to assess participants' reflections on their musical ear. In contrast, by using music perception tasks, the present study bypasses the limitations of using self-assessments as a reliable instrument for evaluating musical and language abilities. This aligns with the observations of Slevc and Miyake (2006), who noted that several studies have failed to establish a clear link between self-ratings of musical ability and language proficiency.

The viewpoint presented by Balčytytė-Kurtinienė (2015) diverges from that of Gralińska-Brawata and Rybińska (2017), focusing particularly on the pedagogical integration of language and music within English lessons and teaching contexts. Balčytytė-Kurtinienė's (2015) study with Lithuanian English learners aged 9-10 underscores the evident advantages for students with heightened musical aptitude when subjected to musical instruction, notably in activating linguistic skills. This phenomenon is linked to the inherent similarities between language and music, hinting at a positive transfer between the two domains. The research further aligns with findings suggesting a substantial relationship between music and the refinement of phonetic capabilities (Balčytytė- Kurtinienė 2015, 417). Additionally, evidence from Balčytytė-Kurtinienė's (2015, 419) short-term experiment supports the notion that musical instruction provides phonetically relevant benefits. Exposure to song-based phonetic instruction positively influenced Lithuanian learners' English skills, enhancing their auditory discrimination of minimal sound pairs, stress recognition, vowel articulation, rhythmic comprehension, and intonation perception.

In contrast to the two previously mentioned studies, the research conducted by Garcia and Schwab (2023) delves into the perception of non-native stress contrasts. Their study examines the impact of native language (Spanish, French, Korean) and musical aptitude on stress detection in Spanish by French and Korean listeners. Their findings highlight significant differences attributed to participants' native languages, suggesting that L1 characteristics, musical aptitude, and proficiency in L2 contribute to the observed variability in non-native stress perception. This underscores the need for further investigation into various language combinations and experimental tasks to elucidate the complexities of stress perception. The experiment conducted in this thesis with Czech participants aligns with this call for expanded exploration in diverse linguistic contexts and experimental paradigms.

1.1.2 Musical aptitude and L2 pronunciation proficiency

Within the realm of linguistic studies, there is a notable exploration of how musical aptitude interacts with L2 pronunciation accuracy, as well as the intriguing connection between perception and production in language learning.

A study by Jekiel and Malarski (2021) involving fifty Polish-speaking adults learning English as an L2 examined the impact of musical aptitude on vowel acquisition. Vowels are often compared to musical notes due to their inherent timbre, pitch, intensity, and duration (Jekiel and Malarski 2021, 1667). Participants underwent pre- and post-training assessments to evaluate changes in vowel production, comparing results with their musical aptitude scores and experience. Pre-existing rhythmic memory was identified as a significant predictor before training while post-training, musical experience notably influenced the production of more native-like English vowels. However, not all vowels were equally affected, with those closest to the model pre-training showing the most improvement. The results indicate that learners' prior proficiency in L2 pronunciation significantly influences their ability to attain a native-like accent. Additionally, there appears to be a potential connection between rhythmic memory and vowel acquisition before training, along with the impact of years of musical experience after training, suggesting that specific musical abilities may aid in acquiring foreign language accents. However, individuals' initial L2 pronunciation skills before accent training remain the primary determinant of their proximity to the model after training (Jekiel and Malarski 2021, 1679).

Another study focusing on the connection between L2 pronunciation accuracy and musical aptitude is by Nardo and Reiterer (2009). They assessed 66 adults using tests, including L2 pronunciation proficiency, a musical aptitude test, and an English grammar assessment. The results revealed strong correlations between musical aptitude, productive phonetic talent, and grammatical sensitivity (237). Specifically, rhythm perception ability showed the most significant correlation with language measures, followed by pitch discrimination and self-evaluated singing scores. The authors concluded that a well- developed rhythm perception ability, coupled with good pitch perception and a liking for singing, are key components for achieving talent and expertise in foreign language pronunciation (238). While Nardo and Reiterer (2009) focused on the correlation between musical aptitude and L2 pronunciation accuracy, particularly highlighting the significance of rhythm perception ability, Llanes-Coromina et al. (2018) explored the practical application of rhythm-based interventions in improving pronunciation skills. The objective was to assess the effectiveness of short-term training with rhythmic beat gestures on the pronunciation of L2 during a reading-aloud task among high school students (Llanes- Coromina et al. 2018). In the study, 59 students were randomly divided into two groups: one group received training with beat gestures, while the other did not. Results indicated that students who received training with beat gestures demonstrated enhanced pronunciation skills, particularly in terms of accentedness, comprehensibility, and fluency, compared to those who did not receive such training. These findings contribute to the growing body of research supporting the utility of beat gestures as effective pedagogical tools in L2 learning contexts.

A study conducted by Pastuszek-Lipińska (2008) aimed to explore the relationship between music education and L2 acquisition, particularly focusing on the perception and production of sounds. The study involved 106 Polish participants, both musicians and non-musicians, who were asked to repeat phrases (shadowing speech) after hearing them multiple times. Their productions were then evaluated by native English speakers, with musicians receiving higher scores and being rated as closer to native speaker production. Pastuszek-Lipińska (2008) interpreted these findings as preliminary evidence suggesting that musicians possess better abilities in perceiving and producing foreign language sounds compared to non-musicians. Therefore, the study indicates that musicians may handle foreign speech material more effectively. These results suggest that musical training could potentially influence L2 acquisition, even after a short period of training. In contrast, Slevc and Miyake (2006) delve deeper into the music-language relationship by examining the correlation between musical ability and L2 proficiency across various domains, including phonological skills, syntax, and lexical knowledge. Their research with 50 adult native speakers of Japanese assessed L2 proficiency across four domains: receptive phonology, productive phonology, syntax, and lexical knowledge, while also considering factors like age of L2 immersion, language use patterns, exposure, and phonological short- term memory. The findings revealed that musical ability predicted L2 phonological proficiency, both receptive and productive, even when accounting for other factors, although it did not explain unique variance in L2 syntax or lexical knowledge.

Intriguing results surfaced from Milovanov's (2010) study, indicating a clear correlation between participants' musical aptitude and their proficiency in L2 pronunciation. Surprisingly, phonemic discrimination ability was found to be unrelated to musical aptitude or pronunciation skills. The research involved 46 Finnish university students, categorized into non-musical individuals, choir members, and English Philology students, who underwent comprehensive testing, including a pronunciation assessment, a phonemic listening discrimination task, and the Seashore test to evaluate musical aptitude. Results revealed that participants with higher musical aptitude exhibited superior English pronunciation skills compared to those with lower musical aptitude, suggesting a significant link between musical and linguistic abilities. Interestingly, while extracurricular music practice positively influenced overall musical aptitude, it did not directly impact pronunciation skills. However, despite not actively practising English phonemes in a classroom setting, choir members demonstrated pronunciation abilities equivalent to English Philology students, indicating that engagement in musical activities might concurrently enhance processes underlying linguistic analysis.

1.1.3 Musical aptitude and L2 perception

Turning now to studies focusing exclusively on sound perception, such as phoneme discrimination and identification of pitch, duration, and overall auditory abilities, a noteworthy investigation by Rodríguez (2021) examines the relationship between language proficiency and specific musical components. Rodríguez conducted assessments with 69 Spanish second-grade students, measuring their musical aptitude across various musical components and their ability to discriminate phonemes in English. The phonological awareness test comprised 12 listening exercises, wherein participants identified the odd-one-out in a group of three words, targeting phonemes challenging to detect in English. However, Rodríguez's study lacks elaboration on the specifics of the English test, which could have provided valuable insights. Despite this limitation, the findings revealed a positive correlation between musical aptitude and the identification of foreign language sounds. Particularly innovative is Rodríguez's exploration of which musical components correlate with different aspects of phonological awareness. Notably, phonological awareness showed the strongest positive correlation with pitch and duration. While Rodríguez anticipated a stronger link with rhythm, she suggests that the limited scope of the phonological awareness test—solely incorporating isolated words—may explain the inconclusive results in this area.

Building upon Milovanov et al. (2010) findings, which suggest that extracurricular music practice positively affects general musical aptitude but not pronunciation skills, this relationship was further explored in the study by Talamini et al. (2018). The aim of Talamini et al. (2018) was to investigate the connection between musical aptitude, training, and L2 English learning among Italian students aged 11 to 15 years old. Specifically, the study focused on assessing perceptual-auditory skills, such as dictation, and non-perceptual skills, like grammar, to elucidate this connection. Participants underwent assessments for musical aptitude using the PROMS test (Law and Zentner 2012), alongside English grammar and dictation tasks. Notably, musicians outperformed non-musicians in the musical aptitude test, raising questions about the extent to which musical aptitude is innate versus influenced by training, as discussed in the introduction. Overall results revealed that music training, rather than musical aptitude, significantly influenced performance in dictation tasks. This suggests that students undergoing music training exhibited superior performance in dictation skills compared to their counterparts, highlighting the potential impact of music training on L2 learning. Talamini et al. (2018, 5) discuss possible explanations for this phenomenon, suggesting that the multisensory integration inherent in learning a musical instrument enhances the ability to focus on auditory features, such as L2 phonology.

Shifting focus to studies centred on languages other than English, Perfors and Ong's (2012) investigation delved into the impact of phonetic training on the perception of phonetic contrasts in non-native languages among both musicians and non-musicians. This research aimed to explore how training influences various types of phoneme contrasts, specifically examining timing-based contrasts like the Hindi contrast and pitch-based contrasts such as the Mandarin tonal contrast. The authors emphasized the challenges adults encounter in improving poor phonetic perception through training. The study involved 96 native English-speaking adults from the University of Adelaide and its surrounding community. The results unveiled that musicians exhibited superior perception for both types of contrasts, regardless of any training effects. This suggests that musicians may possess broader advantages beyond solely pitch perception, although the benefits observed with timing-based contrasts were relatively smaller. It’s plausible that musicians exhibit superior overall auditory processing abilities, commonly referred to as having a “better ear” (Perfors and Ong 2012, 842).

1.1.4 Neurological perspectives on music and language integration

Turning attention to the neuroscientific domain, several studies investigated the correlation between music and language through neural and behavioural measures.

Chobert et al. (2014) conducted a two-year longitudinal study to examine the influence of musical training on linguistic abilities in native language processing at the preattentive level. Elementary school children in the third grade in France, without prior musical experience, were randomly assigned to either music or painting training. Using the mismatch negativity (MMN) as a measure of cortical responses, participants were tested on syllables varying in vowel frequency, duration, and voice onset time (VOT) before training, after 6 months, and after 12 months. Initially, no group differences were detected, but after 12 months, the music group displayed enhanced preattentive processing of syllabic duration and VOT, evident from increased MMN amplitude. These enhancements are likely due to active musical training rather than innate predispositions for music. However, as was observed in previous subchapter research Slevc and Miyake (2006) indicated that both musical aptitudes and musical training can positively impact phonological processing of speech sounds. Therefore, while active musical training may have played a significant role in the observed enhancements, it’s plausible that individual differences in musical aptitude could also contribute to variations in the outcomes. The same research team (François et al. 2013) conducted a comparable longitudinal study, once more enrolling children in either music or painting training. Before training, there were no discernible group differences. However, over subsequent testing sessions, the music group exhibited improved speech segmentation skills, which refer to the ability to extract words from continuous speech, as indicated by both behavioural and electrophysiological measures, whereas the painting group did not show such improvements. These findings underscore the role of music in facilitating speech segmentation, emphasizing its importance for speech perception and language development in children.

Another study focusing on even younger children, aged 5 to 6 years old, was conducted by Hyde et al. (2009), aiming to investigate the impact of instrumental music training on structural brain changes and associated behavioural improvements. By examining whether structural brain differences observed in adult experts stem from innate predispositions or early training, the research utilized deformation-based morphometry (DBM) to analyse brain changes in children undergoing 15 months of instrumental music training compared to a control group. The instrumental group received weekly half-hour private keyboard lessons, while the control group participated in weekly group music classes. Results revealed significant structural brain plasticity in motor and auditory areas after just 15 months of training, correlating with improvements in motor and auditory- musical skills. These findings suggest that training-induced brain plasticity, rather than inherent biological factors, likely underpins structural brain differences observed in adult experts.

Another investigation shedding light on distinctions in the brains of musicians compared to non-musicians was carried out by Marques et al. (2007), focusing on pitch perception. They investigated the impact of musical expertise on the detection of pitch variations in an unfamiliar language. French adults, both musicians and non-musicians, were exposed to sentences spoken in Portuguese, where the final words exhibited prosodic congruity or incongruity through pitch variations. Results revealed that musicians outperformed non-musicians, particularly when the pitch variations were subtle and challenging to detect. Additionally, analysis of event-related brain potentials indicated that musicians exhibited faster processing of prosodic congruity and incongruity compared to non-musicians, with an average difference of 300 milliseconds. These findings suggest that musical expertise enhances the discrimination of pitch, facilitating the processing of pitch variations not only in music but also in language. Furthermore, the study highlights the influence of semantics on the perception of acoustic prosodic cues, underscoring the interconnectedness between music and language processing.

Milovanov et al. (2008) conducted a thorough study to investigate the link between musical aptitude and second language pronunciation skills in elementary school children.

They administered various tests, including a musical aptitude test, listening discrimination tests for phonemes and chords, a production test of English phonemes, and an MMN experiment. Results revealed that children with superior linguistic skills demonstrated better musical abilities. ERP data further supported these findings, showing more pronounced sound-change evoked activation in children with advanced linguistic skills when exposed to music stimuli. These results suggest a correlation between musical and linguistic skills, implying shared neural mechanisms underlying both abilities.

Rossi et al. (2020) conducted a study investigating whether meaning is similarly extracted from spoken and sung sentences. Twenty German native speakers (10 female) participated in the study, with a mean age of 38.65 years (range: 28–53 years). The language material comprised 88 German sentences, evenly split between semantically correct and incorrect. Participants listened to these sentences while performing a correctness judgment task. Employing a multi-methodological approach combining EEG, fNIRS, and behavioural data, the study revealed that the extraction of meaning from sentences is equally processed in both spoken and sung conditions.

In a study by Wong et al. (2007), brainstem encoding of linguistic pitch was investigated in ten amateur musicians and ten non-musicians with no prior exposure to tone languages. The aim was to explore potential differences in neural responses between musicians and non-musicians in processing linguistic pitch patterns. During the experiment, participants watched a video while listening to three randomly presented Mandarin stimuli, which were resynthesized to vary only in fundamental frequency (F0). Results revealed that musicians exhibited more robust and faithful encoding of linguistic pitch compared to non- musicians. These findings not only suggest a common subcortical basis for music and speech processing but also imply a possible reciprocal relationship in corticofugal tuning, providing neurophysiological insights into the enhanced language-learning abilities observed in musicians.

In summary, these studies collectively demonstrate that musical training enhances various aspects of language processing, including phonological discrimination, semantic comprehension, and structural brain changes. Musicians consistently exhibit superior performance in linguistic tasks compared to non-musicians, suggesting shared neural mechanisms underlying musical and linguistic abilities. Musical expertise not only shapes the anatomical structure of the brain but also influences its functional organization. Brain regions implicated in music processing, such as the Heschl’s gyrus, secondary auditory cortex, planum temporale, and anterior part of the superior temporal gyrus (STG), are also crucial for various aspects of language processing (Marques et al. 2007). These shared neural substrates suggest a potential overlap in the cognitive mechanisms underlying both music and language abilities.

1.1.5 The role of music in L2 teaching

In the realm of English Language Teaching (ELT), the integration of music has emerged as a promising pedagogical approach with multifaceted benefits. Scholars have increasingly recognized the inherent connection between music and language, highlighting its potential to enhance language learning outcomes. This subchapter delves into the exploration of music incorporation into L2 teaching, drawing upon seminal studies that illuminate its pedagogical implications and efficacy.

Zybert and Stepień (2009) shed light on the effective integration of music into English language classrooms, emphasizing its potential as a pedagogical tool. They suggest that while music is often relegated to the realm of entertainment in language learning settings, it holds significant untapped potential for enhancing language acquisition. They conducted a study to explore the relationship between musical aptitude and foreign language (FL) learning, particularly focusing on English. Drawing on the theory of multiple intelligences by Gardner, which suggests a connection between linguistic and musical intelligence, the study aimed to assess whether students' musical intelligence influenced their FL learning outcomes. Two groups of 15-16-year-old Polish learners of English, one from a music school and the other from a non-music school, participated in the research. The study included tests to measure music perception, speech perception, music production, and speech production abilities. Results indicated significant correlations between music and language perception and production, suggesting that students' ability to articulate FL sounds accurately was influenced by their musical aptitude. It underscores the importance of considering musical intelligence in the language classroom and suggests that musically talented language learners may benefit from tailored instruction to enhance their learning outcomes.

Transitioning to Fonseca-Mora et al. (2011), their study investigated the potential benefits of incorporating musical elements into FL learning. Recognizing the inherent connection between music and language, the study aimed to explore whether continuous exposure to auditory input through musical exercises would enhance language learning outcomes for all students, regardless of their level of musical development. Forty-nine students aged 11 to 13 participated in the study, all of whom were 6th grade primary students. The study incorporated musical activities, songs, and instrumental music into daily teaching units, with a focus on emphasizing auditory input. Data collection instruments included English proficiency tests, Pimsleur's language aptitude battery, Seashore's musical test, Gardner and Lambert's Attitude and Motivation Test Battery, and a questionnaire to evaluate attitudes towards music activities. Direct observation of lessons was also conducted. The results of the study indicated that integrating musical elements into the language classroom positively impacted students' language learning experiences, fostering a relaxed yet motivating and productive atmosphere while enhancing concentration, creativity, and linguistic memory retention.

Balčytytė-Kurtinienė (2015) investigated the pedagogical benefits of integrating music into ELT to enhance linguistic skills. The study explored how music, especially songs, can break the monotony of lessons and foster collectivism, while also leveraging musical aptitude to activate cognitive skills and improve information retention in language learning. Results revealed that musical instruction, particularly through song, significantly enhanced phonetic skill formation in Lithuanian learners of English, emphasizing heightened auditory discrimination and proficiency in stress patterns, vowel reduction, rhythm, and intonation. These findings underscore the potential of music-based instruction to engage students and enrich their overall linguistic competence.

Lastly, Culp's (2017) study investigated the relationship between phonological awareness and musical aptitude among second-grade students. Using a correlational design, standardized measures such as the Intermediate Measures of Music Audiation (IMMA) and the Phonological Awareness Test 2 (PAT-2) were administered to a sample of students in a rural Pennsylvania charter school. Results revealed a moderate, positive relationship between phonological awareness and tonal musical aptitude, suggesting that students with higher musical aptitude may also demonstrate enhanced phonological skills. These findings have significant implications for music teaching, as understanding students' musical aptitudes can inform tailored instruction to support both musical and phonological development. By incorporating musical aptitude testing into teaching practices, educators can create inclusive learning environments that cater to individual student needs, ultimately fostering greater musical and linguistic achievement.

Overall, the integration of music into ELT offers multifaceted benefits, including improved language proficiency, heightened engagement, and enhanced cognitive skills, making it a valuable pedagogical tool in language education.

1.2 Musical aptitude tests

When exploring the content assessed by musical aptitude tests, they typically target five core categories of abilities (Nardo and Reiterer 2009, 216-217). These include tonal, rhythmic, kinaesthetic, aesthetic, and creative abilities. Tonal abilities encompass skills such as pitch perception, tonality sense, and harmony-polyphony comprehension. Rhythmic abilities involve tasks like abstracting meter, perceiving rhythmic structures, anticipatory rhythm comprehension, tempo synchronization, and tempo regulation. Kinaesthetic abilities come into play during activities like singing or playing instruments, involving skills such as improvising, expressiveness, and auditory recall. The last two categories, aesthetic, and creative abilities are somewhat ambiguous. Aesthetic abilities relate to expression, appreciation, and emotional connection, while creative abilities pertain to originality and flexibility. Additionally, Carroll (1993, 373) proposed a different categorization of auditory perception factors, identifying 31 factors of musical ability divided into four subgroups of discrimination factors. These encompass general sound discrimination (including pitch, timbre, intensity, duration, and rhythm), sound frequency discrimination (involving frequency attributes of tones like melody note changes or chord complexity), sound intensity and duration discrimination, and musical sensitivity and judgmental factors, which involve assessments of musicality in musical passages (Carroll 1993, 373).

Musical aptitude tests evaluate individuals' ability to interpret and synthesize auditory stimuli as music, emphasizing personal inference processes. In contrast, music achievement tests focus on evaluating mastery of musical skills and knowledge, often involving analytical descriptions or notational definitions of finished musical products (Gordon 2007, 53). Every author typically holds a unique perspective on the definition and composition of musical aptitude. Consequently, each musicality test has been developed from distinct viewpoints, often critiquing previous works while striving to integrate, enhance, or refine them. Notable tests from the 20th century include those by Seashore (1960), Wing (1948), Bentley (1966) and Gordon (1990).

Seashore's Measures of Musical Talents (1960), the oldest standardized music test available, was first published in 1919 and revised in 1939 and 1956. This test emphasizes basic sensory capacities using a psychophysical approach, requiring subjects to discriminate certain physical characteristics between pairs of tones and aims and is suitable for 10-16 age. The revised version consists of six subtests, each assessing a specific domain of musical aptitude: pitch, loudness, rhythm, time, timbre, and tonal memory. However, despite its age and widespread use, the Seashore Measures of Musical Talents has faced substantial criticism for its atomistic approach. Critics argue that this test's focus on individual sensory capacities may not adequately predict overall musical ability (Nardo and Reiterer 2009, 222).

Wing (1948) identified several weaknesses in Seashore's (1960) music aptitude test, including a lack of consideration for qualities musicians value, measurement of only one aspect of music, small validation groups, inadequate standardization, limited age range applicability, and neglect of musical training effects. The Wing's test, developed between the late 1930s and 1970, aimed to assess music aptitude independently of musical training. It targets individuals aged 8 to 15 and addresses weaknesses in previous tests, offering higher reliability and validity. It assesses both perceptual and cognitive aspects of music aptitude mainly melody, pitch, loudness, and accent, with varying difficulty levels suitable for different age groups and expertise levels. Additionally, Bentley (1966), the scholar behind the Measures of Musical Ability, incorporates assessments of melody, rhythm, and pitch.

Following Seashore, Gordon emerged as a prominent figure in music aptitude testing, introducing a series of tests designed to assess either developmental or stabilized musical aptitude. The Audie (Gordon 1989b) test aimed at children from 3-4, along with the Primary Measures of Music Audiation - PMMA (Gordon 1979) and the Intermediate Measures of Music Audiation – IMMA (Gordon 1982), all fall under the developmental category of musical aptitude tests. The PMMA is tailored for children aged five to nine, while the IMMA is designed for those aged six to twelve. Subsequently, stabilized music aptitude tests like the Musical Aptitude Profile – MAP (Gordon 1995) for ages ten to fifteen and the Advanced Measures of Music Audiation – AMMA (Gordon 1989a) for ages eleven into adulthood were introduced. Gordon (2007) underscores the importance of MAP and similar tests in education, particularly for students making decisions regarding music studies or instrument learning. Detailed diagnostic analyses become crucial, especially considering that musical aptitude stabilizes around age nine. As highlighted in the subchapter on incorporating music in L2 teaching, students with higher musical aptitude may benefit more from musical instructions and various musical components in the L2 classroom (e.g. Fonseca-Mora et al. 2011; Culp 2017).

Law and Zentner (2012), the developers of the PROMS test utilized in this thesis, have acknowledged limitations, and conducted evaluations on the tests formulated by the four previously mentioned scholars. Firstly, many subtests measured a combination of skills rather than specific ones, leading to ambiguity in performance attribution. Additionally, inconsistencies in stimuli quality, including timing, timbre, and intensity, arose from using human performers in audio recordings, affecting test reliability. Secondly, audio sample sounds often sounded distorted or impure due to recording limitations or material degradation over time. Furthermore, issues with battery design, such as unequal stimuli distribution and varied answer formats across subtests, posed challenges. Response bias and guessing patterns were not adequately controlled, affecting data interpretation. Thirdly, procedures for inferring test validity and reliability were outdated, relying on obsolete indicators of internal consistency and infrequent examination of test-retest reliability. Validation procedures lacked detail, hindering robust inferences about test validity. Lastly, crucial aspects of music perception, such as timbre, tuning, and tempo, were not adequately assessed in older musical aptitude tests.

Now, in examining the tests of the 21st century, contemporary music assessment batteries adhere to robust principles of test design and validation. According to Law and Zentner (2012, 3) "[…] these batteries were specifically devised to capture deficits rather than individual differences in musical perception skills within the normal range." For example, the Montreal Battery Evaluation of Amusia – MBEA (Peretz et al. 2003) targets the assessment of amusia, while the Clinical Assessment of Music Perception – CAMP (Kang et al. 2009) focuses on evaluating music perception among adults with cochlear implants. On the other hand, the Musical Ear Test – MET (Wallentin et al. 2010) solely measures melody and rhythm perception skills. The Goldsmith Musical Sophistication Index (Müllensiefen et al. 2012) presents a comprehensive tool for gauging musical abilities in the general population, yet its findings remain preliminary.

This thesis employs the PROMS test developed by Law and Zentner (2012) intending to address a gap in musical ability assessments. Their objective was to create a battery that satisfied four essential criteria: suitability for individuals with diverse musical backgrounds, inclusivity in terms of tested musical perceptual components, specificity in assessing each component, and adherence to contemporary standards of test construction regarding validity and reliability.

1.3 English stress

1.3.1 Production and perception

In examining the intricate nature of stress in English, it's crucial to explore the multifaceted dimensions that underpin this linguistic phenomenon. One perspective involves examining the actions of the speaker when producing stressed syllables, while another perspective involves identifying the acoustic features that lead a listener to perceive a syllable as stressed. These acoustic features, also known as phonetic cues or acoustic correlates of English stress, play a significant role in this thesis's investigation. Stress in speech is a suprasegmental attribute that encompasses entire syllables rather than individual vowels or consonants. A stressed syllable is articulated with increased energy compared to an unstressed one, making it more noticeable within the rhythm of speech (Ladefoged and Johnson 2011).

Stress production usually entails intensified activity in the respiratory muscles, leading to greater volume. Consequently, the speaker expels more air from the lungs, potentially accompanied by heightened laryngeal and articulatory activity. Additionally, there's a magnification of consonant and vowel attributes, like vowel elevation and stop aspiration. Finally, there's an accentuation of pitch, resulting in lower pitches becoming lower and higher pitches becoming higher (Ladefoged and Johnson 2011).

Transitioning from stress production to perception, attention shifts to how listeners perceive stress in speech. According to Ladefoged and Johnson (2011, 112) when we are listening to other people, "[…] we are probably putting together all the cues available in a particular utterance in order to deduce the motor activity […] we would use to produce those same stresses […]." In further exploration, various scholars have identified a range of cues contributing to stress perception, which will be examined in subsequent discussions.

1.3.2 Acoustic correlates of stress

Roach (1998) asserts that recognizing stress in syllables depends on their prominence, which is shaped by four key perceptual factors: loudness, length, pitch, and vowel quality. These correspond to the acoustic properties of intensity, duration, fundamental frequency (F0), and formant structure, respectively. Loudness is paramount, with stressed syllables typically perceived in English as louder. However, altering loudness alone may not significantly impact perception if other attributes remain constant. Syllable duration also contributes to perceived stress, with stressed syllables having a longer duration than their unstressed counterparts, which is a cross-linguistic correlate of stress (Ortega-Llebaria and Prieto 2011), although in Czech duration is reserved for marking phonemic vowel length differences and does not cue stress (e.g. Volín and Weingartová 2014). F0 variations, reflecting vocal fold vibrations, can create noticeable differences between syllables, affecting their perceived stress (Roach 1998, 74) . Furthermore, since stressed syllables in English are associated with pitch accents—fluctuations in F0 that shape sentence intonation—English listeners use the temporal relationship between syllables and pitch accents to perceive stress prominence (Ortega-Llebaria and Prieto 2011). Additionally, vowel quality within syllables can influence their prominence, particularly when contrasting with weaker syllables (Roach 1998, 74). Roach (1998) explains that a syllable is likely to be perceived as prominent if it contains a vowel that is qualitatively different from the neighbouring vowels. This distinct vowel quality can make the syllable stand out as stressed against the backdrop of more frequently encountered vowels in weak syllables. Although English listeners consider these factors, vowel duration in itself does not constitute a cue to stress prominence. These factors interact to create prominence in stressed syllables, with pitch and length (corresponding to F0 and duration) exerting the strongest influence (Roach 1998, 74).

The investigation aims to explore whether listeners from different linguistic backgrounds perceive stress differently, examining the influence of their L1 and the roles of pitch and length, as studied by Lehiste and Fox (1992). Languages can be classified into stress, tone, and quantity based on their suprasegmental properties, wherein intensity, F0, and duration play distinct roles. Consequently, individuals speaking languages of different suprasegmental types may interpret suprasegmental cues differently. For example, Estonian, characterized as a quantity language, contrasts with English, classified as a stress language. Native Estonian and English speakers were tested using speech and nonspeech signals to assess responses to variations in duration and amplitude. English speakers tended to prioritize amplitude cues, while Estonian speakers favoured duration cues. The absence of F0 variation had a greater impact on speech-like stimuli perception than on noise stimuli, underscoring F0's significance in stress perception. Hence, the reliance on cues in stress perception appears to be language-specific, suggesting that "[…] a listener's native language may affect perceptual judgment of phonetic segments […]" (Lehiste and Fox 1992, 419).

Another acoustic property that influences stress perception is spectral tilt. Ortega- Llebaria and Prieto (2011) delved into this cue in their study, examining its role alongside duration and intensity as stress correlates in Catalan and Spanish. They controlled for formant frequency differences between vowels in stressed and unstressed environments. Traditionally, vowels are acoustically characterized by their formant frequencies (Ladefoged and Johnson 2011, 196). In English, unstressed vowels undergo qualitative reduction as they centralize towards the schwa sound. Ortega-Llebaria and Prieto (2011) highlighted the challenge of conflicting results regarding spectral tilt as a stress correlate, possibly because of vowel reduction. However, their findings showed that changes in spectral tilt were significant only in Catalan and specific vowels, suggesting a connection with formant frequency differences rather than stress contrast. Additionally, they found that duration consistently indicated stress across all vowels in both languages, while the significance of intensity was less pronounced. Ortega-Llebaria and Prieto (2011) concur with Roach (1998) regarding the reliability of F0 and duration as stress correlates. Furthermore, they explored the relationship between stressed and unstressed syllables and stress correlates, revealing that in stress-accented languages like English, F0 variations align with stressed syllables, underscoring the importance of the timing of F0 changes in stress perception. However, stress contrast persisted even in unaccented sentences, suggesting a reciprocal association between F0 and other stress indicators. Moreover, Eriksson and Heldner (2015) conducted a study to analyse and model the acoustics of word stress, revealing that for English data, duration and spectral emphasis emerged as the most robust indicators of stress level. While F0 level also showed a significant correlation, its influence was not as pronounced. Their analysis encompassed unstressed, secondary, and primary stress, as well as different speech styles such as spontaneous, phrase, and word. Interestingly, their findings diverged from previous research in terms of the degree of explained variance, with spectral emphasis (17,5%), duration (14,2%), and F0 level (7,3%) exhibiting varying levels of importance.

Klatt (1976) investigated the role of duration as a stress indicator by analysing the literature on segmental duration. He correlated this with perceptual studies on duration discrimination and psychophysical research on listeners' capacity to make linguistic judgments solely based on duration cues. If an alteration in duration exceeds the threshold for just-noticeable differences in segmental duration, it holds the capacity to convey valuable information from the speaker to the listener. Klatt (1976) concluded that in English, duration often acts as a primary perceptual cue for distinguishing between various linguistic elements, including: inherently long and short vowels, voiced and voiceless fricatives, phrase-final and non-final syllables, voiced and voiceless postvocalic consonants (evidenced by changes in the duration of preceding vowels in phrase-final positions), stressed and unstressed or reduced vowels, and the presence or absence of emphasis. Thus, Klatt (1976, 1219) asserts that "an important perceptual corelate of an unstressed or reduced vowel is a reduction in its duration."

The phonetic cues to stress include intensity, duration, F0, and formant structure, with intensity and duration identified as primary indicators. Spectral tilt also emerges as a significant cue, particularly in certain vowels. These cues interact to create prominence in stressed syllables, contributing to cross-linguistic variations in stress perception.

1.3.3 English versus Czech stress

When comparing Czech and English, one notable difference lies in their approach to stress. English is characterized as a stress-timed language, where variable word stress dominates rhythmic timing (Ladefoged and Johnson 2011). Additionally, stress serves multiple purposes in English. Firstly, it is utilized within sentences to provide emphasis to particular words or to contrast one word with another. Secondly, stress signals the syntactic category of words, distinguishing between noun-verb pairs such as an ˈinsult (noun) and to inˈsult (verb). Similarly, stress operates within compounds, as seen in a ˈwalkout (compound) versus to ˈwalk ˈout (verb). Lastly, variations in stress can be linked to the grammatical structure of words, as in ˈdiplomat, dipˈlomacy, and diploˈmatic (Ladefoged and Johnson 2011).

In contrast, Czech follows a fixed-stress pattern, typically emphasizing the first syllable of a word, although there are exceptions such as certain foreign words like "ahoj," "pardon," where stress is optionally on the second syllable, or prepositional phrases such as "na zahradě" that are treated as single phonological words and the preposition thus takes the stress (if it is syllabic) (Palková 1994). Czech stress has no contrastive function at the level of individual words, and it is not influenced by vowel quality, length, or word morphology (Skarnitzl and Eriksson 2017). Duběda and Votrubec (2005) conducted research on the Czech language, revealing that the most reliable predictor of stress, either individually or when considered alongside other factors, is F0. While duration's utility is somewhat limited, it operates independently of stress and serves a phonological function in distinguishing between short and long vowels. In Czech, the syllable immediately preceding the accented one primarily governs most durational, dynamic, and intonational contrasts. According to findings from Skarnitzl and Eriksson's study (2017), unlike other languages investigated, Czech does not exhibit distinct acoustic cues on the stressed syllable to indicate lexical stress. In Czech, the stressed syllable does not demonstrate heightened prominence in terms of pitch variation, duration, sound pressure level (SPL), or spectral emphasis compared to unstressed syllables (Skarnitzl and Eriksson 2017). Some minimal tendencies suggest a delayed rise, with slightly elevated acoustic parameters observed on the second syllable following the stress. This absence of acoustic marking on the stressed syllable aligns with the non-contrastive nature of lexical stress in Czech. Consequently, Czech learners frequently encounter difficulties in both producing and recognizing English stress due to Czech's limited phonetic implementation of stress. This phenomenon has been characterized as "stress deafness" (Skarnitzl and Rumlová 2019).

Volín and Weingartová (2014) investigated how British and Czech speakers of L2 English utilize acoustic cues for stress in speech production. Sixteen participants were tasked with reading English BBC news bulletins, and their speech was analysed for vowel duration, F0, SPL, and spectral slope. Additionally, a perception test was conducted, where Czech listeners rated the native likeness of utterances from other participants. Their findings revealed that Czech speakers, when speaking L2 English, employ significantly fewer acoustic indicators of stress compared to their British counterparts, which aligns with linguistic observations detailed by Skarnitzl and Eriksson (2017) and Duběda and Votrubec (2005) regarding the stress of Czech language. In the context of English, the most notable differences were observed in F0, with Czech speakers exhibiting lower F0 in stressed vowels, and in SPL, where they showed minimal difference between stressed and unstressed syllables. However, in terms of duration, Czech speakers demonstrated almost native-like patterns in distinguishing between stressed and unstressed vowels. This may be attributed to the Czech language's use of vowel duration to convey phonological quantity differences between long and short vocalic phonemes, suggesting a transfer of sensitivity to duration cues to a different function in a foreign language.

In summary, when Czech speakers are producing English, the disparity in expressions of word stress can be categorized by their relative significance of SPL, fundamental frequency, and spectral slope, although to a lesser extent. Conversely, duration differences show minimal variation between both groups, indicating that Czech speakers readily acquire native-like patterns in this aspect when speaking English.

1.4 Research questions and hypotheses

Based on the insights from the literature examined in the preceding sections, the following research questions (Q_X) then their operationalized versions showing the specific measures (Q_{XB}) and their corresponding hypotheses (H_X) have been formulated.

Q_1: Is there a relationship between L2 English proficiency and the ability to perceive L2 English stress?

Q_{1B}: Do the participants' higher LexTALE scores correlate with a higher probability to correctly discriminate the stress placement in English-pronounced nonsense syllable strings?

H_1: LexTALE scores will be positively associated with L2 stress discrimination, such that higher English L2 proficiency will coincide with a greater success at correctly discriminating L2 English stress.

Q_2: Is there a relationship between musical aptitude among Czech learners of L2 English and the ability to perceive L2 English stress?

Q_{2B}: Do the participants' higher PROMS-S scores correlate with higher probability to correctly discriminate the stress placement in English-pronounced nonsense syllable strings?

H_2: PROMS-S scores will be positively associated with L2 stress discrimination, such that higher PROMS-S scores will coincide with greater success at correctly discriminating L2 English stress. This hypothesis is based on previous research by Gralińska-Brawata and Rybińska (2017), Balčytytė-Kurtinienė (2015), and Garcia and Schwab (2023), which found that higher musical aptitude relates to better stress perception. Consequently, a similar relationship is expected to be observed in this thesis.

Q_3: Which specific musical perception skills (tonal, spectral/qualitative, temporal, dynamic) within musical aptitude are related to the perception of L2 English stress?

Q_{3B}: Which musical domains assessed by the PROMS-S test, i.e. tonal (melody, pitch), spectral/qualitative (timbre, tuning), temporal (rhythm, rhythm-to-melody, tempo), and dynamic (beat), correlate with the higher probability to correctly discriminate the stress placement in English-pronounced nonsense syllable strings?

H_3: In alignment with the literature, it is hypothesized that certain musical domains will be positively associated with L2 stress discrimination. Specifically, higher scores in the tonal (melody, pitch) and temporal (rhythm, rhythm-to-melody, tempo) domains will coincide with greater success at correctly discriminating L2 English stress. Notably, no studies have directly examined the relationship between stress perception and specific musical domains. However, this hypothesis is supported by Culp (2017), who found a relationship between phonological awareness and the tonal domain; Rodríguez (2021), who found that phonological awareness (specifically the ability to discriminate phonemes) could be predicted by pitch and duration; and Nardo and Reiterer (2009), who identified rhythm and pitch as significant correlates of expertise in FL pronunciation, including L2 pronunciation. Therefore, a similar relationship is expected to be observed in this study.

2 Method

This chapter presents the experiment aimed at Czech speakers to investigate the relationship between their capacity to discern stress in English as their L2, their musical aptitude, and their proficiency in English, measured by vocabulary size. It outlines the participants involved, the procedural steps undertaken, the tests administered, and the methods employed for data analysis.

2.1 Participants

54 adults, all native Czech speakers with normal hearing abilities and no attentional or neurological disorders, participated in the experiment. The participants were students of a bachelor's degree at Palacký University in Olomouc either in the English Philology (N = 40) or the Musicology (N = 14) program at the time of the experiment. The sample included both female (N = 45) and male (N = 9) participants with an average age of 19.87 (min = 19, max = 24).

2.2 Procedure

The experiment started by inviting students to participate during their lessons, highlighting the chance to receive scores indicating both their musical aptitude and proficiency in recognizing English stress. Students were then asked to fill out a Google form indicating their availability for testing, their field of study and their native language (other than Czech speakers were not included in the statistics). The testing sessions took place in a computer classroom at the Faculty of Arts, Palacký University in Olomouc, throughout December 2023. Participants attended two separate sessions within a two-week period, during which they completed either the music aptitude test or the stress discrimination task along with a Lexical Test for Advanced Learners of English (LexTALE) to assess their proficiency (Lemhöfer & Broersma 2012). Each computer was supplied with a reliable pair of headphones to ensure participants experienced quiet surroundings and optimal volume levels.

At the beginning of each session, the administrator (author of this thesis) delivered instructions in Czech to all participants collectively to ensure clarity, as the instructions integrated into the tests were in English, and not all participants had the same level of English proficiency. For the musical aptitude test PROMS-S (Law and Zentner 2012), participants accessed a shared link and followed the instructions independently at their own pace, with the test lasting around 40 minutes. Participants were instructed to remain seated and await a signal from the administrator, which was displayed on the board, indicating that all participants had completed the test. This was done to ensure a calm testing environment for everyone. Before the English stress discrimination test, participants completed the LexTALE test (Lemhöfer & Broersma 2012). At the start of the stress discrimination task, participants submitted their LexTALE score, years of English learning, age, gender, and assigned ID for anonymity. Following this, they proceeded independently, adhering to instructions that included a practice round, with the test typically lasting around 30 minutes. Prior to participation, all individuals signed an informed consent form (refer to Appendix A: Informed Consent for the exact wording).

2.3 Tests

2.3.1 Musical aptitude test

Musical aptitude was assessed using the Profile of Music Perception Skills (PROMS), which offers various testing options based on the number of domains covered. These include the FULL-PROMS (9 subtests), Micro-PROMS (10 minutes), and mini-PROMS (4 subtests). For this thesis, the PROMS-S, comprising 8 subtests, was deemed the most appropriate choice. The battery of AAX discrimination tasks evaluated learners' perceptual musical skills across 8 domains, which can be categorized as tonal (melody, pitch), spectral/qualitative (timbre, tuning), temporal (rhythm, rhythm-to-melody, tempo), and dynamic (beat) (Law and Zentner, 2012).

Now, the specifics of these domains within the PROMS-S, as described by Law and Zentner (2012), will be examined. In the melody trials, which feature a sequence of musical tones or pitches arranged coherently, participants encounter monophonic sequences of eighth notes characterized by a consistent rhythm. Rhythm, defined as a pattern of sounds and silences organized temporally, comprises straightforward patterns characterized by consistent note intensities. The rhythm-to-melody domain challenges listeners to identify rhythmic patterns when they are embedded within a melody, requiring them to focus on the rhythmic structure of the melody independently of its pitch contour. The beat or accent, linked to musical meter and speech stress, evaluates the ability to perceive the relative emphasis placed on specific notes within a rhythmic pattern. Regarding tempo, the pace of music performance, it is measured in beats per minute (bpm), with variations in the standard stimuli ranging from 7 bpm (easy) to 1 bpm (difficult). In the pitch domain, which pertains to the perceived frequency of sound, a 2,000-ms sinusoidal waveform with a 250-ms linear onset and offset ramp was utilized to reduce emphasis on the beginning and end points, ensuring consistent intensity levels. Difficulty levels were adjusted by varying the pitch distinction between standard and comparison stimuli, ranging from 7 to 50 cents (equivalent to 2 to 12 Hz), centred around 440 Hz, a common frequency in music serving as the standard tuning pitch for musical instruments, potentially providing familiarity for classically trained musicians. Timbre, referring to the unique character of a sound, was explored using original instrument sounds within chords to create a diverse range of timbral qualities. In tuning, the process of aligning a musical instrument's pitch to a standard frequency, a C chord (comprising C, E, and G notes) was employed, with particular focus on manipulating the E note to produce an intentionally dissonant sound.

The tasks for melody, beat, and timbre consist of 10 trials each, while pitch, tuning, rhythm, rhythm-to-melody, and tempo tasks include 8 trials each. Every trial commenced with a musical excerpt (A), followed by its repetition (A), and then a subsequent sound (X) that could be either identical to or different from the first sound (A). Participants were asked to rate their confidence in comparing the sounds using the following scale: "definitely same," "probably same," "I don't know," "probably different," and "definitely different". The Department of Psychology at the University of Innsbruck provided access to the PROMS-S for the purpose of administering the test in this experiment and analysing the results. Initially, the scoring system was as follows: 1 point for a correct response with a "definite" rating, 0.5 points for a correct response with a "probable" rating, and 0 points for an incorrect response with "probable or definite" rating or an "I don't know" rating. However, to align the scoring with the statistical objectives and to better differentiate between varying degrees of error, the scoring system was modified as follows: 2 points for a correct response with a "definite" rating, 1 point for a correct response with a "probable" rating, 0 points for an "I don't know" rating, -1 point for a "probable" rating of an incorrect response, and -2 points for a "definite" rating of an incorrect response. Additionally, to ensure equitable weight across all domains, the scoring was adjusted to maintain consistent scoring proportions: the difference between the score and the minimum score was divided by the maximum range of the scale within each domain.

2.3.2 Stress discrimination task

The English test, from Tlolková's master's thesis (2018) involving Czech and Spanish learners, was based on AXB discrimination task, implemented using a script for the Demo Window in the Praat program (Boersma and Weenink 1992-2024). The stimuli were 12 CVCVCV pseudowords recorded by 3 native English speakers. The stimuli were combined into 144 AXB trials, in half of which X had the same stress placement as B, and in the other half as A, e. g. [təˈnɛɹiː], [dəˈkɔɪliː], [ˈsiːməlaɪ]. Participants were instructed to focus on the stress patterns of non-existent words. They were informed that they would hear three non-existent English words and were tasked with determining whether the stress pattern of the second word (referred to as X) was more similar to the preceding word A or the following word B. Upon hearing a beep sound, they were prompted to press either button A or B, indicating whether X resembled the preceding or following word, respectively. For information regarding the elicitation of the three-syllable pseudowords and their syllabic structure, please refer to Tlolková's master thesis (2018).

This thesis extends the information provided in Tlolková's thesis (2018) by assessing the acoustic cues to English stress (specifically the F0, intensity, duration, vowel formants, and spectral tilt) within the recorded stimuli to make sure a natural variability was present, and the test did not result in a ceiling effect. To measure the acoustic cues in the stimuli, firstly the syllable boundaries and vowel boundaries were labelled manually (by the author of the present thesis) following the guidelines of phonetic segmentation defined by Machač and Skarnitzl (2009) and then a script in the Praat program (Boersma and Weenink 1992- 2024) was used to extract the values of each measured acoustic correlate. The measured values are reported below, and they were also submitted to regression modelling using R (R Core Team 2024, Wickham 2016, Lüdecke 2018). The plots were generated in R using the package *ggplot2* (Wickham 2016).

	Syllable					
	1		2		3	
	stressed	unstressed	stressed	unstressed	stressed	unstressed
Syllable duration (ms, log-transformed)	5.11 (0.45)	4.81 (0.49)	5.63 (0.26)	5.08 (0.26)	6.05 (0.23)	5.64 (0.29)
Vowel duration (ms, log-transformed)	4.52 (0.31)	4.21 (0.44)	4.96 (0.43)	4.06 (0.47)	5.79 (0.27)	5.33 (0.33)
Syllable intensity (dB)	74.28 (2.09)	69.42 (2.39)	72.43 (1.07)	68.40 (3.10)	71.36 (1.02)	66.45 (2.57)
Vowel intensity (dB)	75.33 (1.04)	70.24 (2.27)	74.25 (1.11)	70.69 (1.89)	71.08 (1.42)	64.30 (3.10)
Syllable F0 (semitones re 100 Hz)	5.97 (4.29)	0.71 (2.13)	4.00 (4.35)	1.55 (2.66)	2.07 (3.96)	-1.45 (3.24)
Vowel F0 (semitones re 100 Hz)	6.09 (4.26)	0.69 (2.12)	5.21 (5.19)	1.28 (2.80)	1.76 (4.46)	-2.19 (3.39)
Spectral tilt (F1 amplitude – F2 amplitude)	12.20 (9.63)	14.70 (7.41)	16.02 (6.94)	11.35 (5.79)	14.03 (6.64)	8.97 (7.18)
F1 (Bark)	4.50 (0.98)	3.99 (0.60)	4.25 (0.71)	4.27 (0.49)	4.74 (1.42)	4.99 (1.02)
F2 (Bark)	11.86 (1.12)	11.50 (1.09)	11.38 (1.37)	10.84 (0.71)	12.36 (1.20)	11.99 (1.14)
F3 (Bark)	14.62 (0.59)	14.52 (0.59)	14.41 (0.66)	14.35 (0.75)	15.25 (0.44)	14.86 (0.38)

Table 1: Means and Standard Deviations in parentheses of acoustic correlates of stress in the English stimuli split by syllable.

The values displayed in Table 1 represent the mean values for individual syllables and measured variables, accompanied by their standard deviations. However, it is important to not only analyse the differences between these means but also examine the distribution of all the measured data. In the following section, a detailed description of each measured stress correlate is provided, along with the visualization of raw data and the presentation of linear regression models for each measured variable.

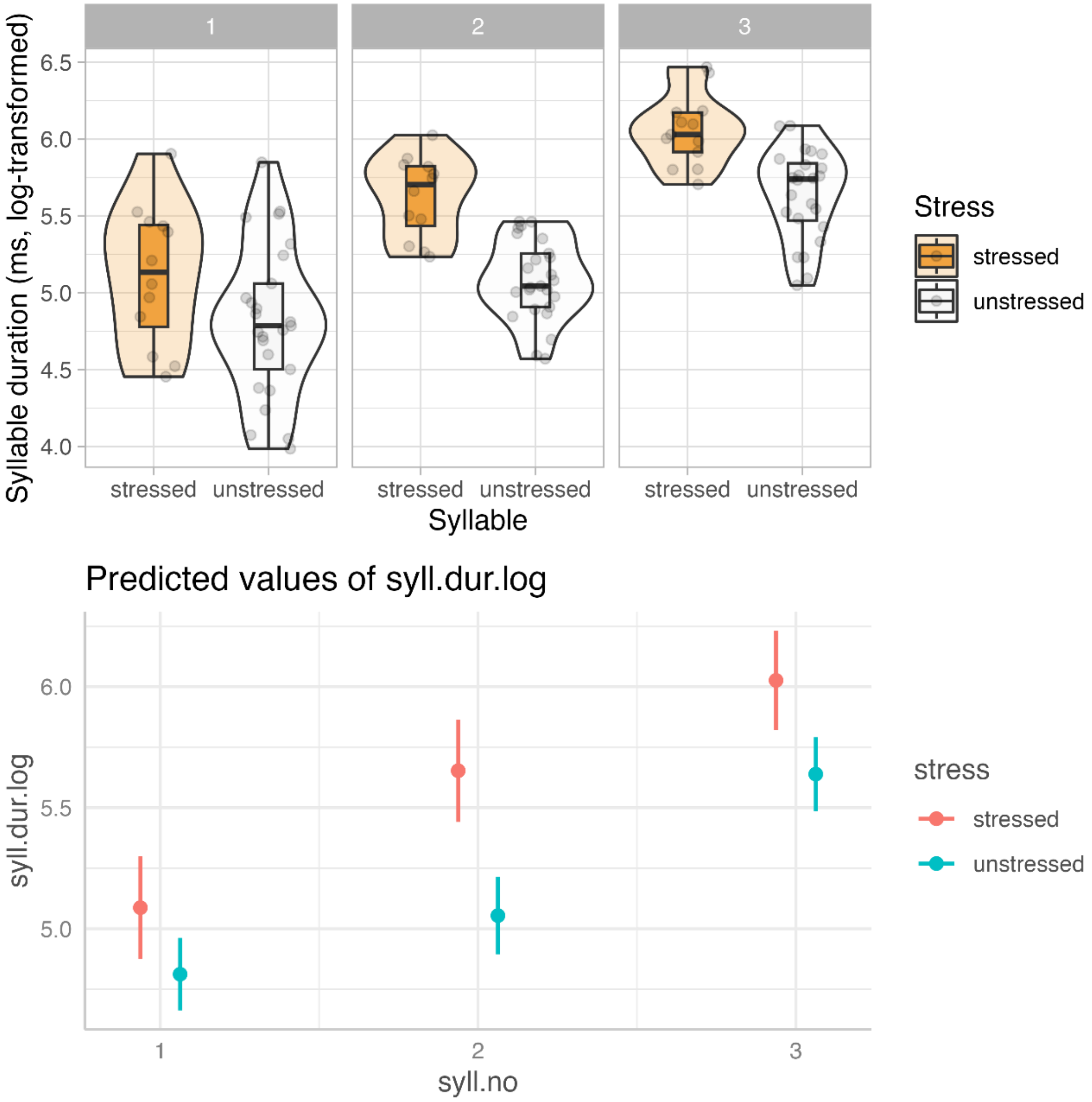


Figure 1: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw syllable durations (in ms, log-transformed) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

The violin plots in Figure 1 (top) clearly suggest that, overall, the first syllable of the stimuli tended to be shorter, and the third syllable longer, than the second. This can be interpreted in terms of word-final lengthening (Bell et al. 2003). Importantly, in all three syllable positions, stress tended to lengthen syllable duration. The syllable duration data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded) and Word as the random effect with varying intercepts. Table 2 gives the model coefficient estimates and Figure 1 (bottom) shows the model- predicted syllable duration values. The model confirmed that Syllable number affected syllable duration (est.: -0.429, SE = 0.049, t = -8.758, $p <$ 0.0001), and so did Stress (est.: 0.21, SE = 0.034, t = 6.12, $p <$ 0.0001).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	5.379	0.05	6.886	108.197	0
stress1	0.21	0.034	98.424	6.12	0
syll.no1	-0.429	0.049	100.89	-8.758	0
syll.no2	-0.025	0.05	104.592	-0.503	0.616
stress1:syll.no1	-0.073	0.051	103.243	-1.436	0.154
stress1:syll.no2	0.089	0.056	50.044	1.586	0.119

Table 2: Coefficients estimated by a mixed-effects linear regression model fitted to the syllable duration data.

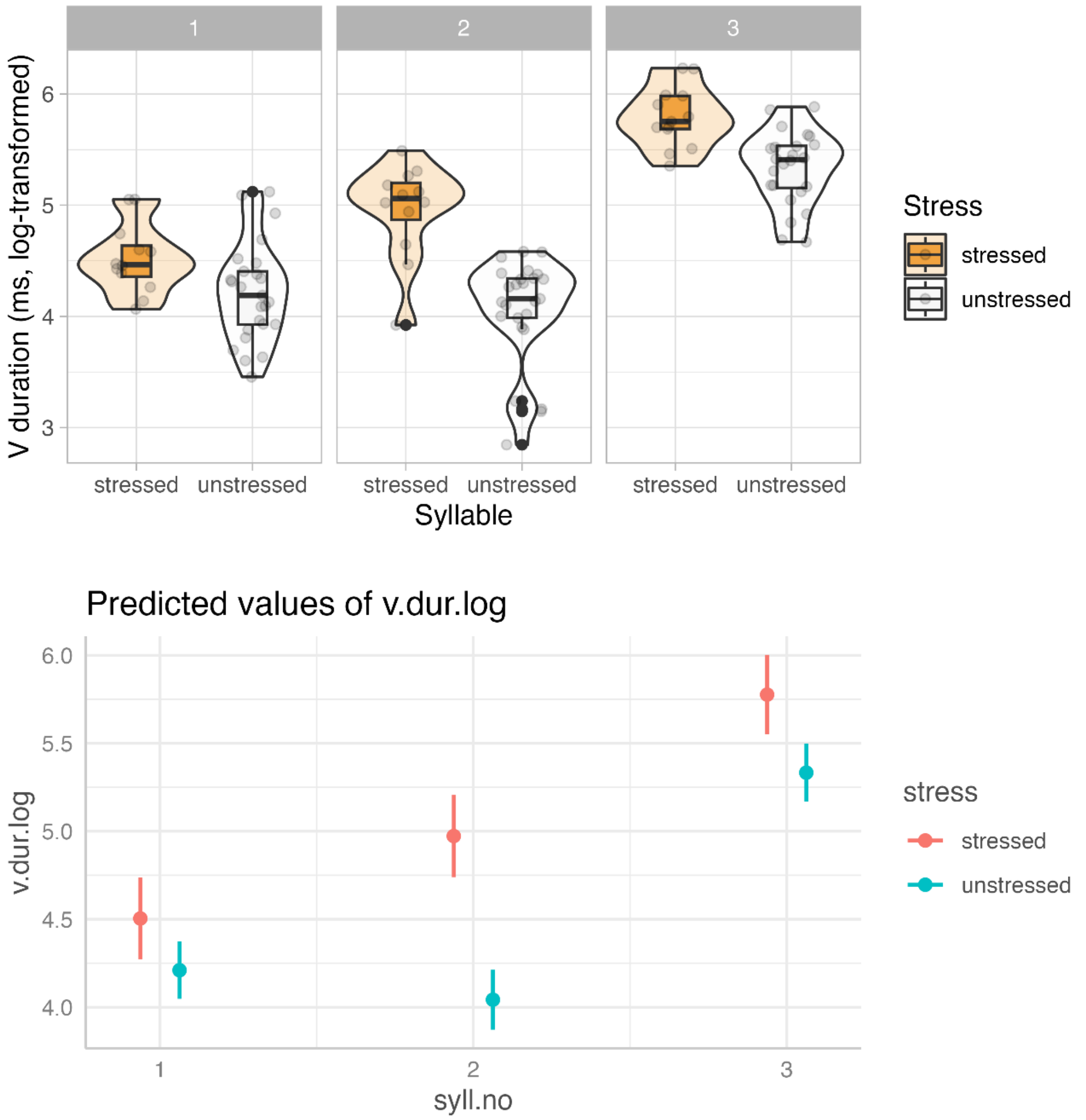


Figure 2: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw vowel durations (in ms, log-transformed) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

Relatedly, the violin plots in Figure 2 (top) clearly suggest that the first vowel of the stimuli tended to be shorter, and the third vowel longer, than the second. The vowel duration data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded) and Word as the random effect with varying intercepts. Table 3 gives the model coefficient estimates and Figure 2 (bottom) shows the model-predicted vowel duration values. The model confirmed that Stress affected vowel duration (est.: 0.278, SE = 0.039, t = 7.077, $p < 0.0001$).

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	4.807	0.049	6.373	97.215	0
stress1	0.278	0.039	98.139	7.077	0
syll.no1	-0.449	0.056	100.869	-8.037	0
syll.no2	-0.299	0.056	104.763	-5.301	0
stress1:syll.no1	-0.131	0.057	101.981	-2.293	0.024
stress1:syll.no2	0.187	0.061	38.763	3.056	0.004

Table 3: Coefficients estimated by a mixed-effects linear regression model fitted to the vowel duration data.

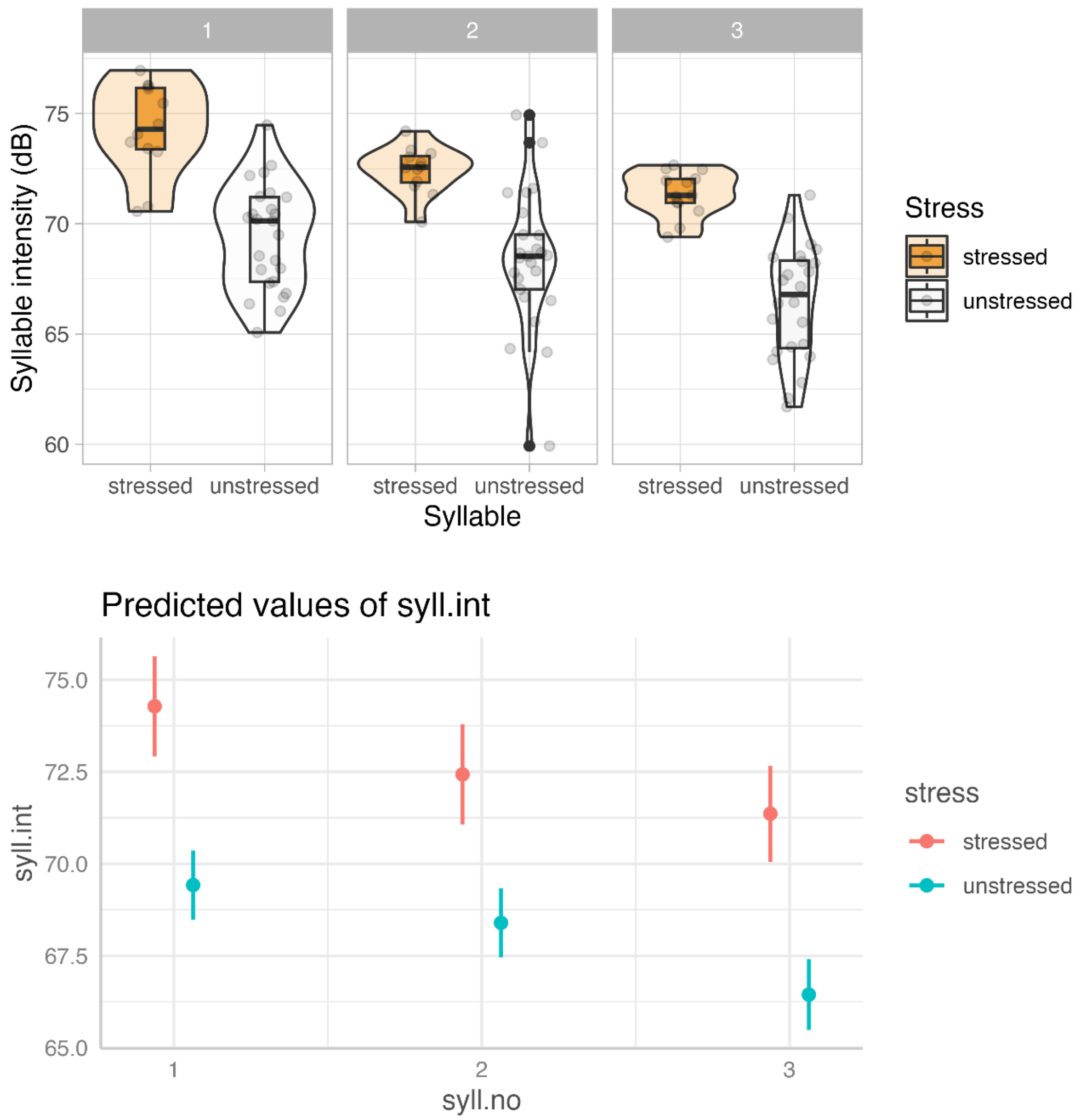


Figure 3: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw syllable intensity (in dB) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

Next, the violin plots in Figure 3 (top) clearly suggest that in all syllable positions stress increased syllable intensity. The syllable intensity data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum- coded).[2] Table 4 gives the model coefficient estimates and Figure 3 (bottom) shows the model-predicted syllable intensity values. The model confirmed that Stress affected syllable intensity (est.: 2.301, SE = 0.239, t = 9.618, p < 0.0001).

[2] Word was not included as a random effect in this model, as a model including it did not converge.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	70.389	0.239	294.187	0
stress1	2.301	0.239	9.618	0
syll.no1	1.462	0.339	4.306	0
syll.no2	0.024	0.339	0.07	0.944
stress1:syll.no1	0.128	0.339	0.378	0.706
stress1:syll.no2	-0.284	0.339	-0.837	0.405

Table 4: Coefficients estimated by a mixed-effects linear regression model fitted to the syllable intensity data.

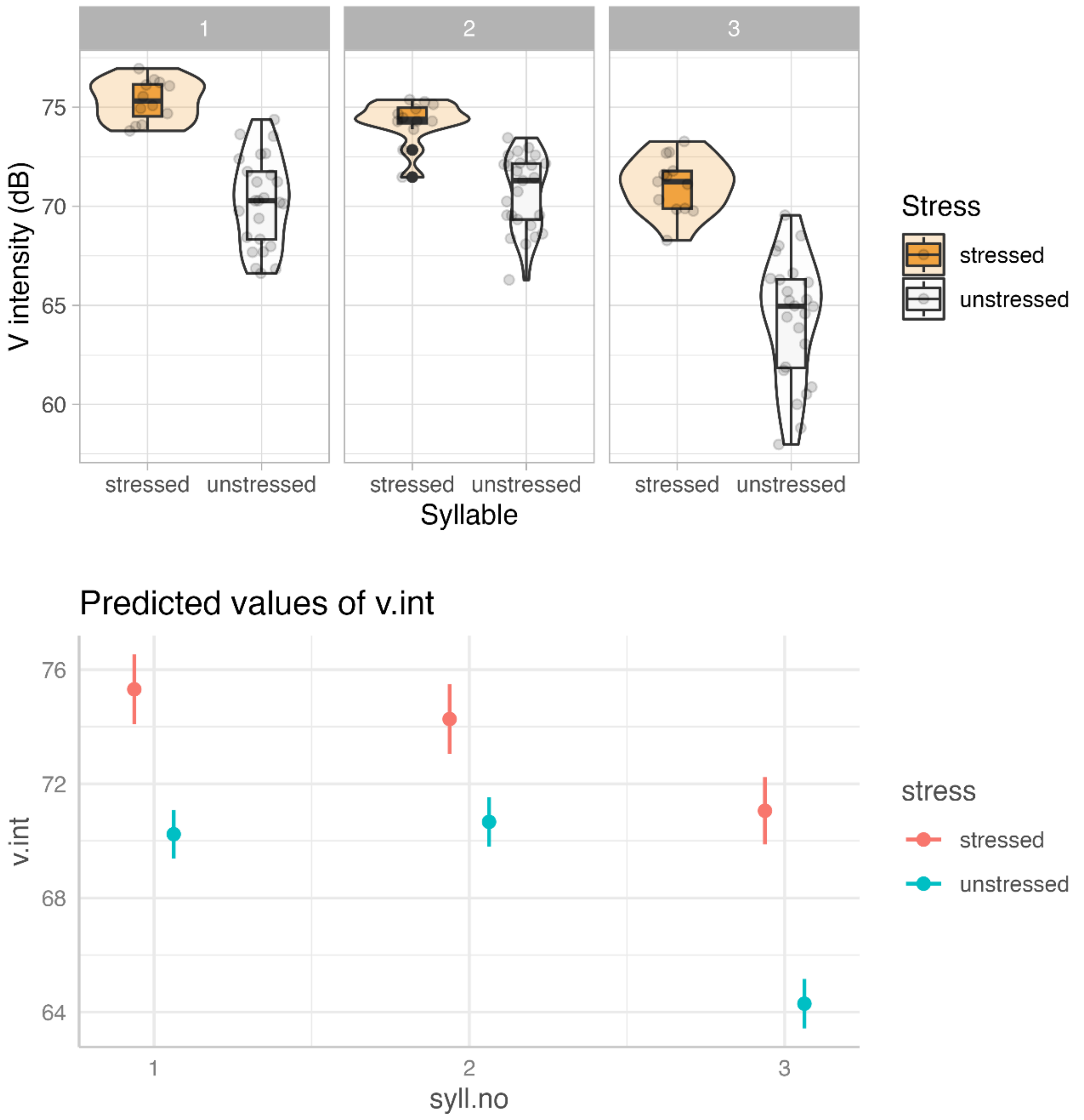


Figure 4: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw vowel intensity (in dB) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

The violin plots depicted in Figure 4 (top) then prominently illustrate that the stressed vowel in all syllable positions exhibited the higher intensity an unstressed vowel, with the most notable contrast between stressed and unstressed vowels observed in the third position. The vowel intensity data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded) and Word as the random effect with varying intercepts. Table 5 gives the model coefficient estimates and Figure 4 (bottom) shows the model-predicted vowel intensity values. The model confirmed that Stress affected vowel intensity (est.: 2.573, $SE = 0.215$, $t = 11.98$, $p < 0.0001$).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	70.972	0.229	4.404	309.901	0
stress1	2.573	0.215	95.523	11.98	0
syll.no1	1.801	0.305	99.283	5.905	0
syll.no2	1.495	0.306	104.744	4.89	0
stress1:syll.no1	-0.035	0.307	99.749	-0.114	0.909
stress1:syll.no2	-0.772	0.313	23.826	-2.464	0.021

Table 5: Coefficients estimated by a mixed-effects linear regression model fitted to the vowel intensity data.

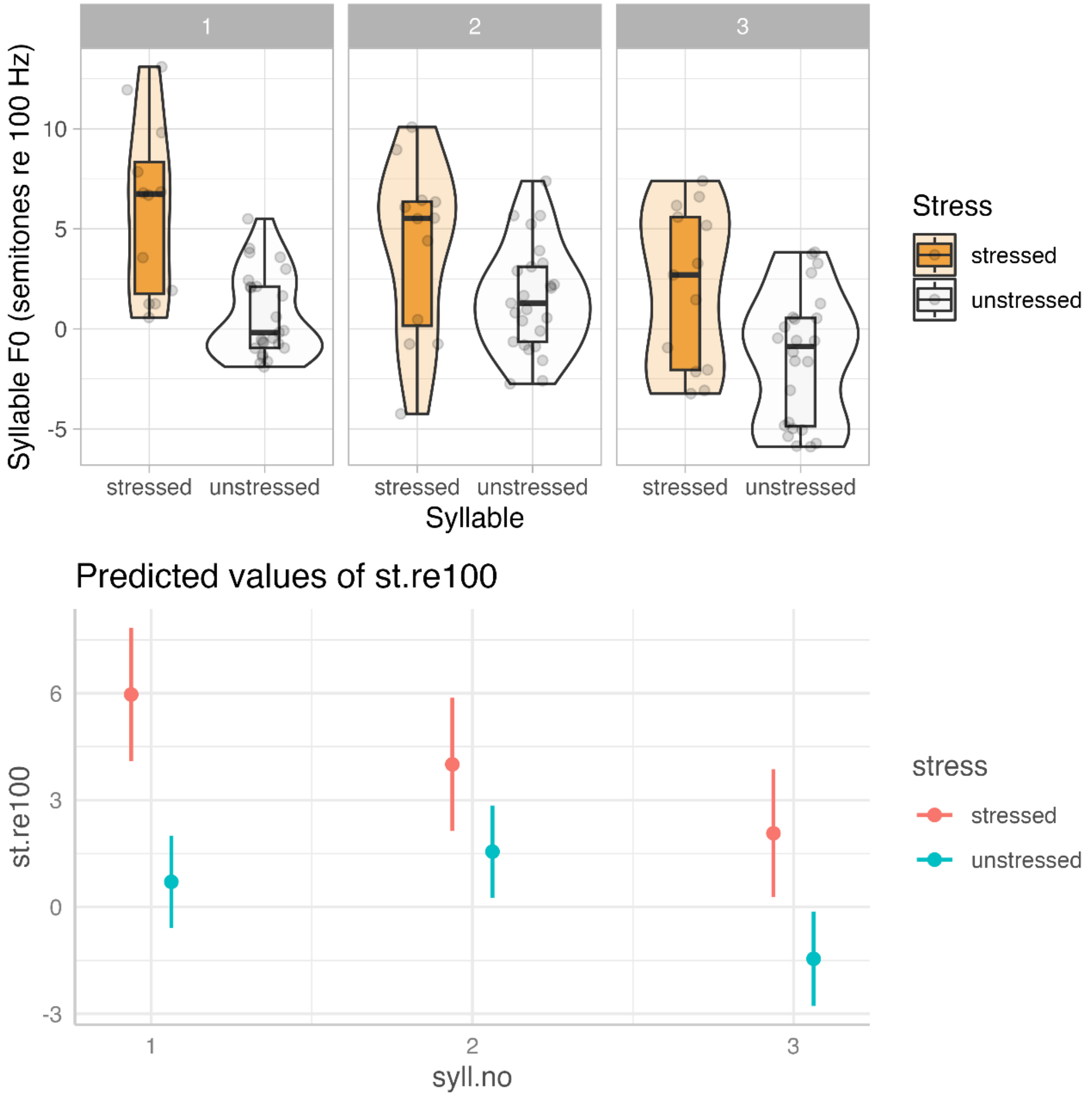


Figure 5: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw syllable F0 (in semitones re 100 Hz) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

The violin plots depicted in Figure 5 (top) suggest that, for all syllable numbers, when a syllable was stressed, it had a somewhat higher F0 than when unstressed. The syllable F0 data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded).[3] Table 6 gives the model coefficient estimates and Figure 5 (bottom) shows the model-predicted syllable F0 values. The model confirmed that Stress affected syllable F0 (est.: 1.872, $SE = 0.329$, $t = 5.691$, $p < 0.0001$).

[3] Word was not included as a random effect in this model, as a model including it did not converge.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.141	0.329	6.507	0
stress1	1.872	0.329	5.691	0
syll.no1	1.195	0.467	2.561	0.012
syll.no2	0.638	0.467	1.367	0.175
stress1:syll.no1	0.757	0.467	1.622	0.108
stress1:syll.no2	-0.647	0.467	-1.386	0.169

Table 6: Coefficients estimated by a mixed-effects linear regression model fitted to the syllable F0 data.

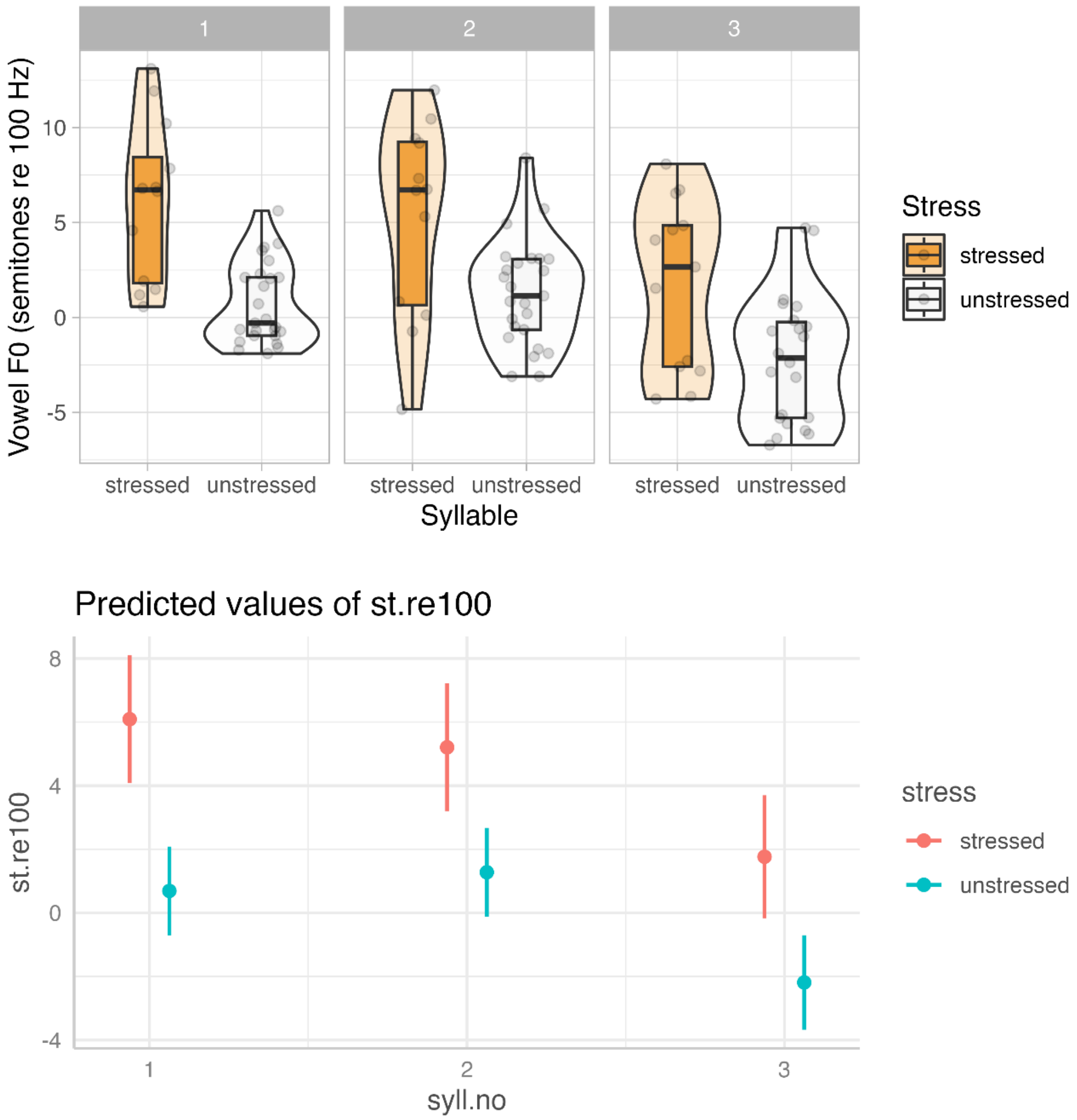


Figure 6: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw vowel F0 (in semitones re 100 Hz) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

Similarly, the violin plots given in Figure 6 (top) suggest that when a vowel was stressed it tended to have a higher F0 than when unstressed, irrespective of whether it was inside the first, second or third syllable of the stimulus. The vowel F0 data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded).[4] Table 7 gives the model coefficient estimates and Figure 6 (bottom) shows the model-predicted vowel F0 values. The model confirmed that Stress affected vowel F0 (est.: 2.214, $SE = 0.356$, $t = 6.221$, $p < 0.0001$).

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	2.139	0.356	6.011	0
stress1	2.214	0.356	6.221	0
syll.no1	1.251	0.504	2.483	0.015
syll.no2	1.102	0.504	2.188	0.031
stress1:syll.no1	0.487	0.504	0.966	0.336
stress1:syll.no2	-0.25	0.504	-0.496	0.621

Table 7: Coefficients estimated by a mixed-effects linear regression model fitted to the vowel F0 data.

[4] Word was not included as a random effect in this model, as a model including it did not converge.

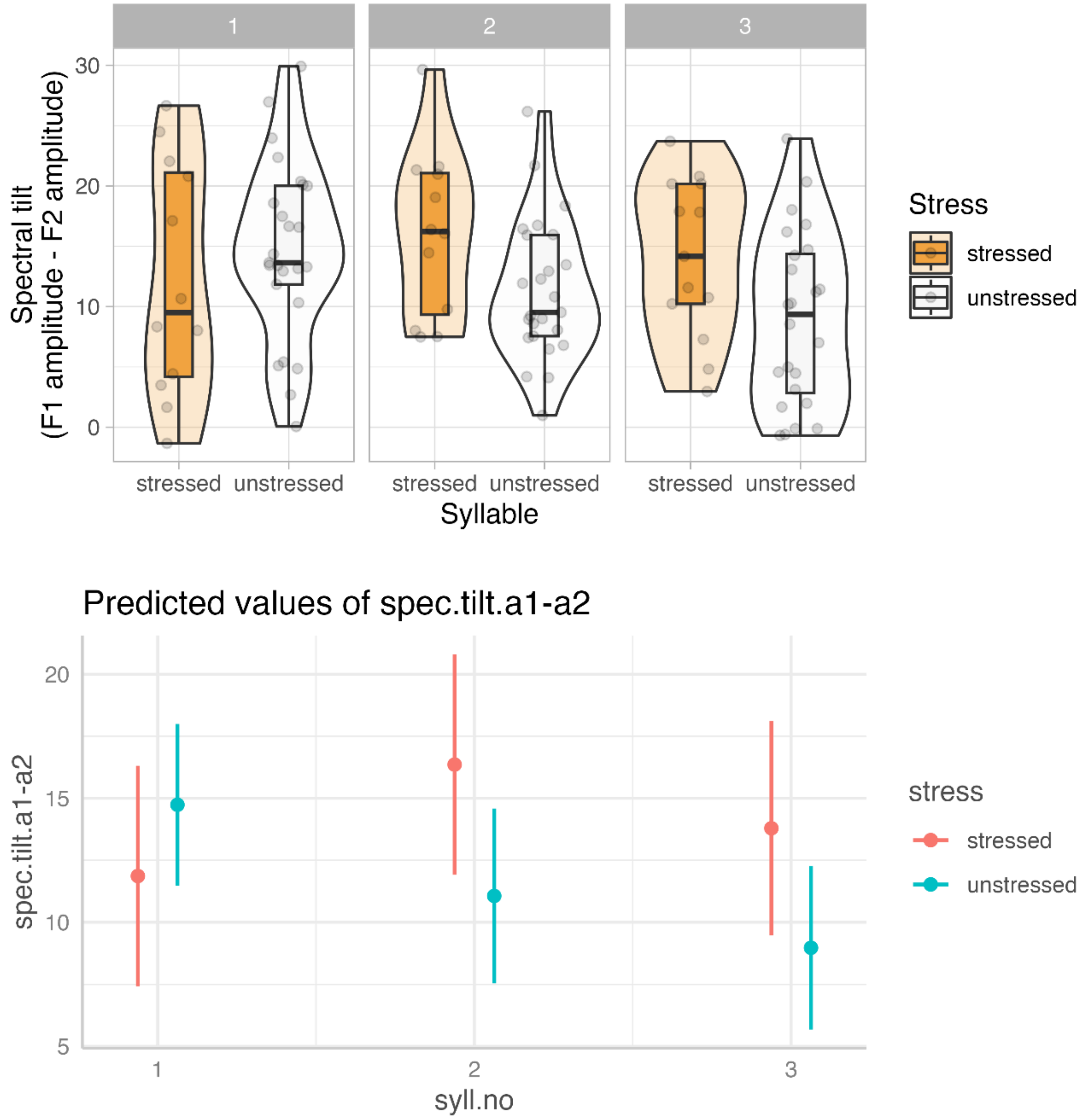


Figure 7: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw spectral tilt (in amplitude of F1 – F2 amplitude) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

In contrast to the consistent effect of stress observed with the acoustic cues described thus far, the violin plots in Figure 7 (top) showing the spectral tilt (measured as the difference of the amplitude of the second formant from the amplitude of the first formant of each vowel) reveal a different pattern. Specifically, across the vowels in first syllables of the stimuli, there is a slight decrease of spectral tilt when stressed compared to when unstressed. In contrast, vowels in the second and third syllables seem to exhibit a somewhat higher spectral tilt when stressed. This suggests that spectral tilt alone may not be a reliable acoustic cue to stress in Tlolková's (2018) stimuli. Still, the spectral tilt data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded) and Word as the random effect with varying intercepts. Table 8 gives the model coefficient estimates and Figure 7 (bottom) shows the model-predicted spectral tilt values. The analysis did not reveal a unidirectional effect of stress.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	12.798	1.225	7.567	10.443	0
stress1	1.207	0.667	98.737	1.808	0.074
syll.no1	0.503	0.955	100.679	0.527	0.599
syll.no2	0.912	0.975	103.971	0.935	0.352
stress1:syll.no1	-2.643	1.005	104.794	-2.63	0.01
stress1:syll.no2	1.44	1.168	73.399	1.233	0.221

Table 8: Coefficients estimated by a mixed-effects linear regression model fitted to the spectral tilt data.

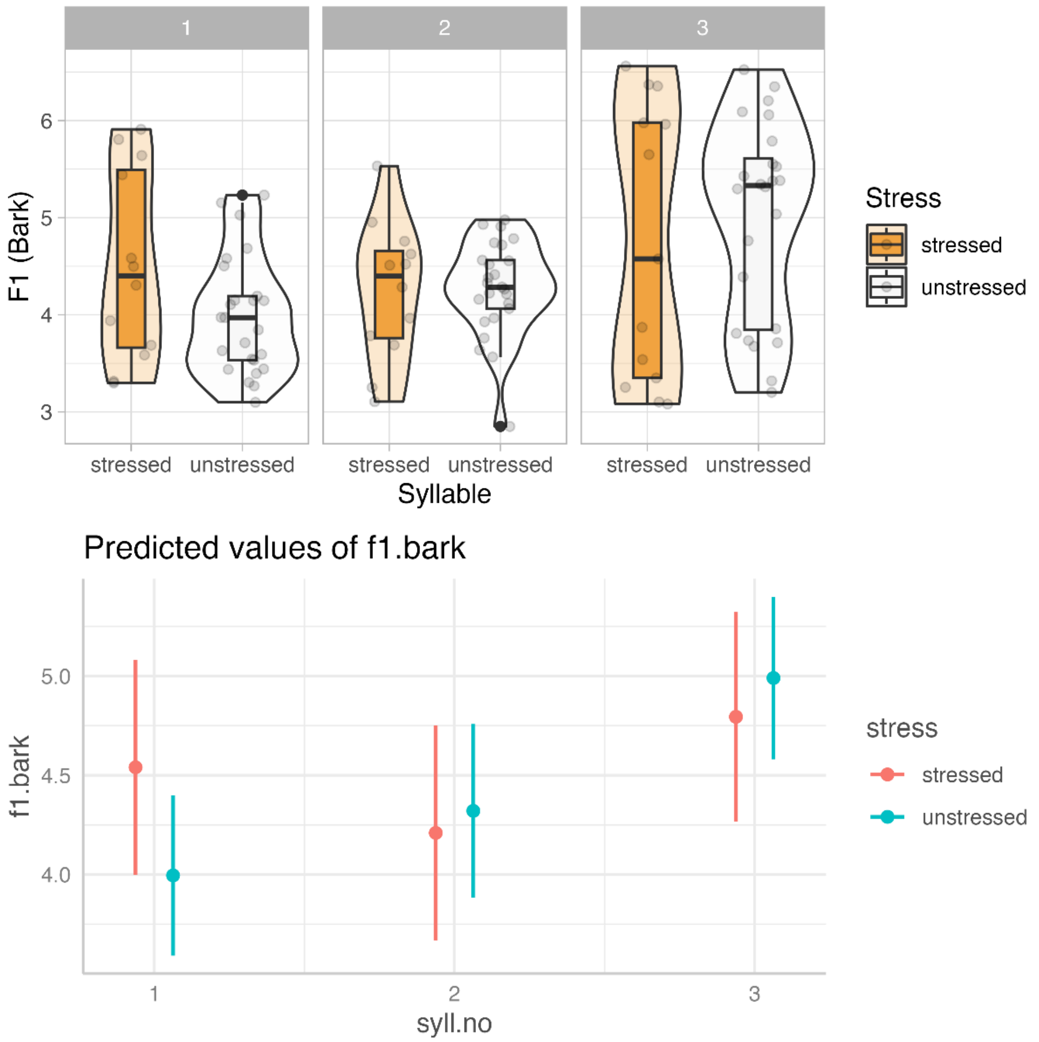


Figure 8: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw F1 (in Bark) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

Similarly to the plots of spectral tilt, the violin plots depicted in Figure 8 (top), showing the F1 values of the vowels in the stimuli, do not reveal a clear and straightforward effect of stress. In the first syllable, the F1 values of vowels under stress are somewhat higher than those of unstressed vowels. Conversely, in the second and third syllables, the values overlap considerably. The F1 data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded) and Word as the random effect with varying intercepts. Table 9 gives the model coefficient estimates and Figure 8 (bottom) shows the model-predicted F1 values. The model did not identify a reliable influence of stress on F1.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	4.475	0.158	7.834	28.249	0
stress1	0.04	0.079	98.882	0.505	0.615
syll.no1	-0.207	0.113	100.601	-1.838	0.069
syll.no2	-0.21	0.115	103.645	-1.818	0.072
stress1:syll.no1	0.233	0.119	104.988	1.95	0.054
stress1:syll.no2	-0.095	0.141	82.263	-0.677	0.5

Table 9: Coefficients estimated by a mixed-effects linear regression model fitted to the F1 data.

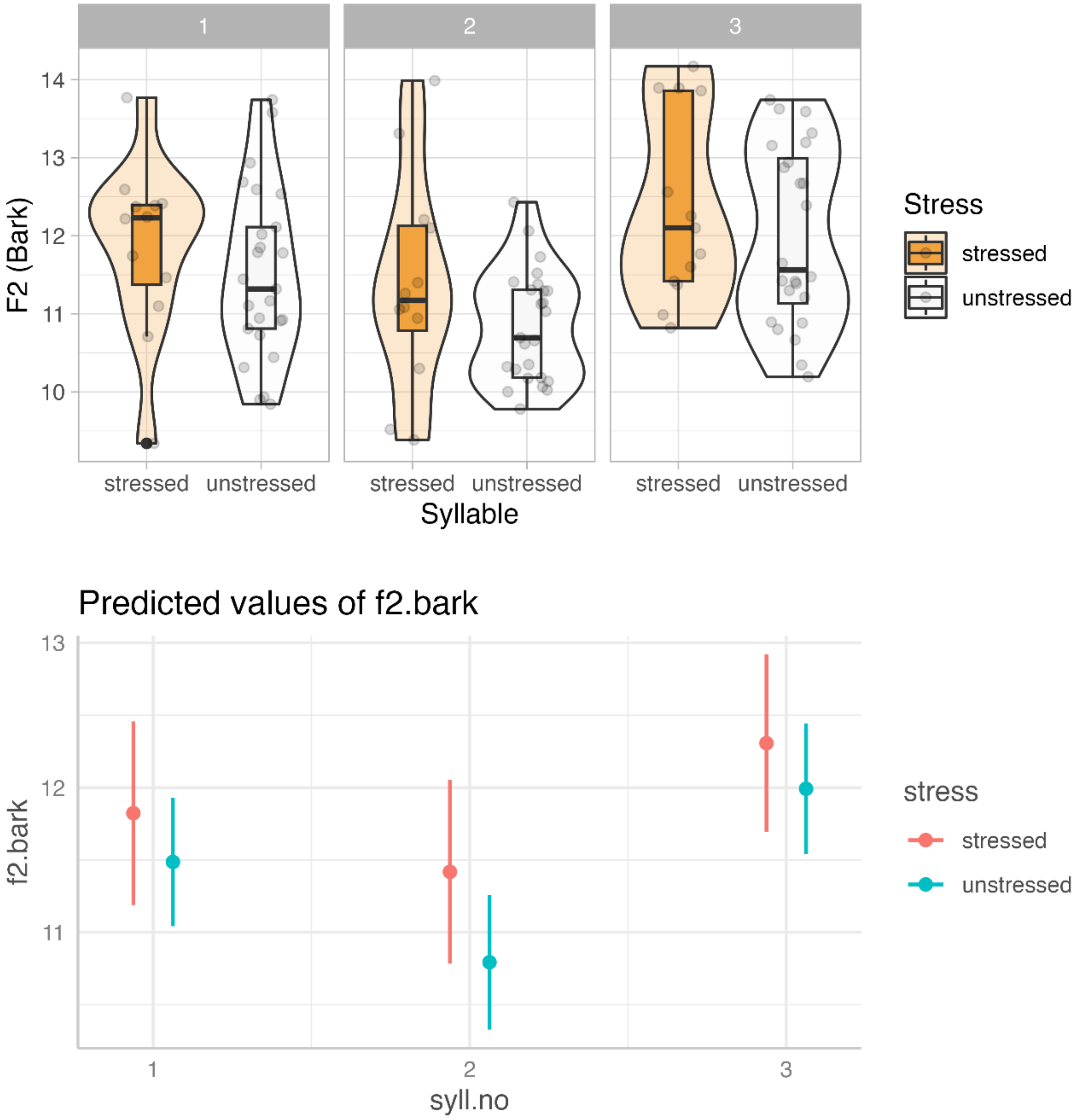
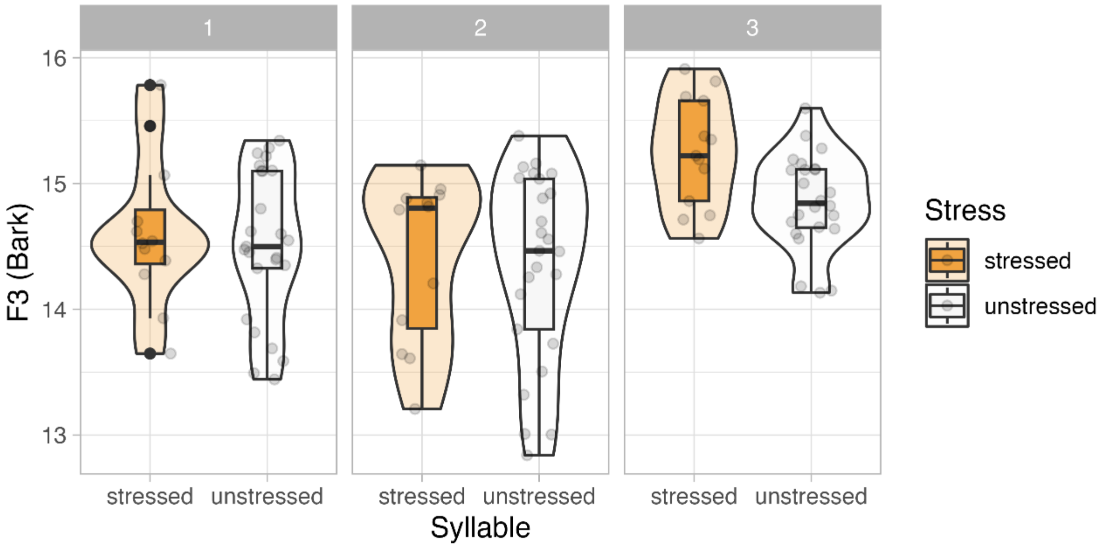


Figure 9: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw F2 (in Bark) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

Next, the violin plots presented in Figure 9 (top) suggest that vowels in stressed syllables tended to have slightly higher F2 values than those in unstressed syllables. The F2 data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded) and Word as the random effect with varying intercepts. Table 10 gives the model coefficient estimates and Figure 9 (bottom) shows the model- predicted F2 values. The model confirmed that Stress had effect on F2 (est.: 0.213, $SE = 0.107$, $t = 1.995$, $p < 0.0001$).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	11.637	0.136	6.764	85.794	0
stress1	0.213	0.107	98.517	1.995	0.049
syll.no1	0.018	0.152	101.094	0.12	0.905
syll.no2	-0.531	0.153	104.768	-3.467	0.001
stress1:syll.no1	-0.044	0.155	102.23	-0.286	0.775
stress1:syll.no2	0.1	0.167	40.931	0.597	0.554

Table 10: Coefficients estimated by a mixed-effects linear regression model fitted to the F2 data.



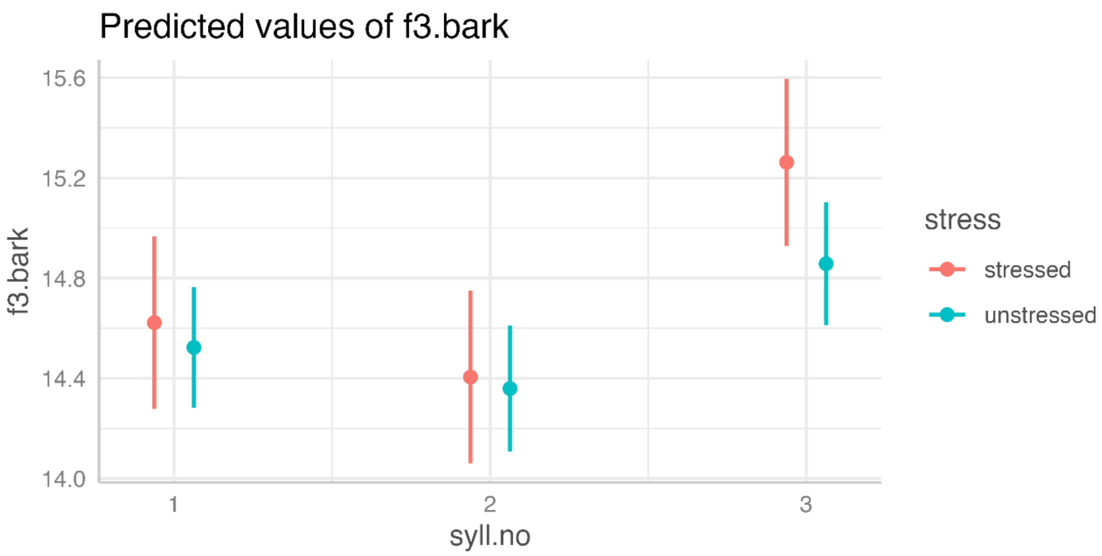


Figure 10: The top panel gives violin plots with overlaid boxplots and jitter plots showing the raw F3 (in Bark) split by syllable number and stress. The jitter plots display the individual measured values. The bottom panel shows values fitted by a linear regression model.

Finally, Figure 10 (top) given the violin plots of the F3 values measured in the vowels of the stimuli. Similarly to F1, no clear influence of stress on the F3 can be observed. The F3 data were fitted to a mixed-effects linear regression model with Stress and Syllable number as the fixed effects (both sum-coded) and Word as the random effect with varying intercepts. Table 11 gives the model coefficient estimates and Figure 10 (bottom) shows the model-predicted F3 values. As expected, the model did not reveal a reliable impact of stress on F3, indicating that F3 alone was probably not a consistent cue to stress within the stimulus set.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	14.672	0.073	9.167	200.611	0
stress1	0.092	0.058	100.264	1.579	0.117
syll.no1	-0.099	0.083	102.172	-1.197	0.234
syll.no2	-0.289	0.083	104.839	-3.471	0.001
stress1:syll.no1	-0.042	0.084	102.944	-0.498	0.619
stress1:syll.no2	-0.069	0.091	48.78	-0.759	0.452

Table 11: Coefficients estimated by a mixed-effects linear regression model fitted to the F3 data.

To summarize, in the present thesis Tlolková's (2018) stimuli were measured for the acoustic properties relevant to stress perception. First, for all the acoustic correlates measured overlaps between the values of stressed and unstressed syllables were observed, indicating that even for experienced listeners it might be difficult to distinguish stressed and unstressed syllables in the stimuli auditorily. However, clear effects of stress on syllable duration, vowel duration, syllable intensity, vowel intensity, vowel F0, syllable F0, and F2 were found, showing that these properties could serve as reliable cues to stress perception within the stimulus set. This was not true of spectral tilt, F1, and F3, whose associations with stress were either absent or complex, suggesting that these three acoustic properties could only potentially contribute to perceived stress in combination with other acoustic cues.

2.3.3 LexTALE

The Lexical Test for Advanced Learners of English (LexTALE) assesses vocabulary knowledge through a lexical decision task, which estimates the size of learners' vocabulary. Participants view sixty lexical items individually and determine whether they are legitimate English words. While primarily measuring lexical knowledge, it has been found to correlate with overall language proficiency in English and surpass self-rated proficiency assessments (Lemhöfer and Broersma, 2012). Furthermore, it is efficiently administered, typically requiring only about 5 minutes to complete. Participants receive scores ranging from 0 to 100 upon completion. Scores between 80 and 100 generally align with CEFR levels C1 and C2, while scores between 60 and 80 are estimated to correspond to level B2. Scores below 59 are anticipated to correlate with proficiency levels of B1 and lower.

2.4 Data analysis

The segmentation, annotation and measurements of the recorded stimuli were performed in Praat (Boersma and Weenink 1992-2024) as presented in section 2. 3. 2. The lexical decision data, musical perception data, and the English stress perception data were all submitted to regression modelling using R (R Core Team 2024) using the packages *lme4* (Douglas et al. 2015) and *ggeffects* (Lüdecke 2018). The plots were generated in R using the package *ggplot2* (Wickham 2016). The complete R script is included in an electronic appendix within the folder titled "Results, R script."

3 Results

3.1 Raw data and their discussion

First, the resulting data are presented in the form of plots and graphs. Since some of these data are aggregated across participants and words, interpretation should be approached with caution. Therefore, the analysis is supplemented by statistical modelling described below in Chapter 3.2. Nevertheless, a preliminary evaluation of the research questions is possible just by observing the data visualizations presented in this section.

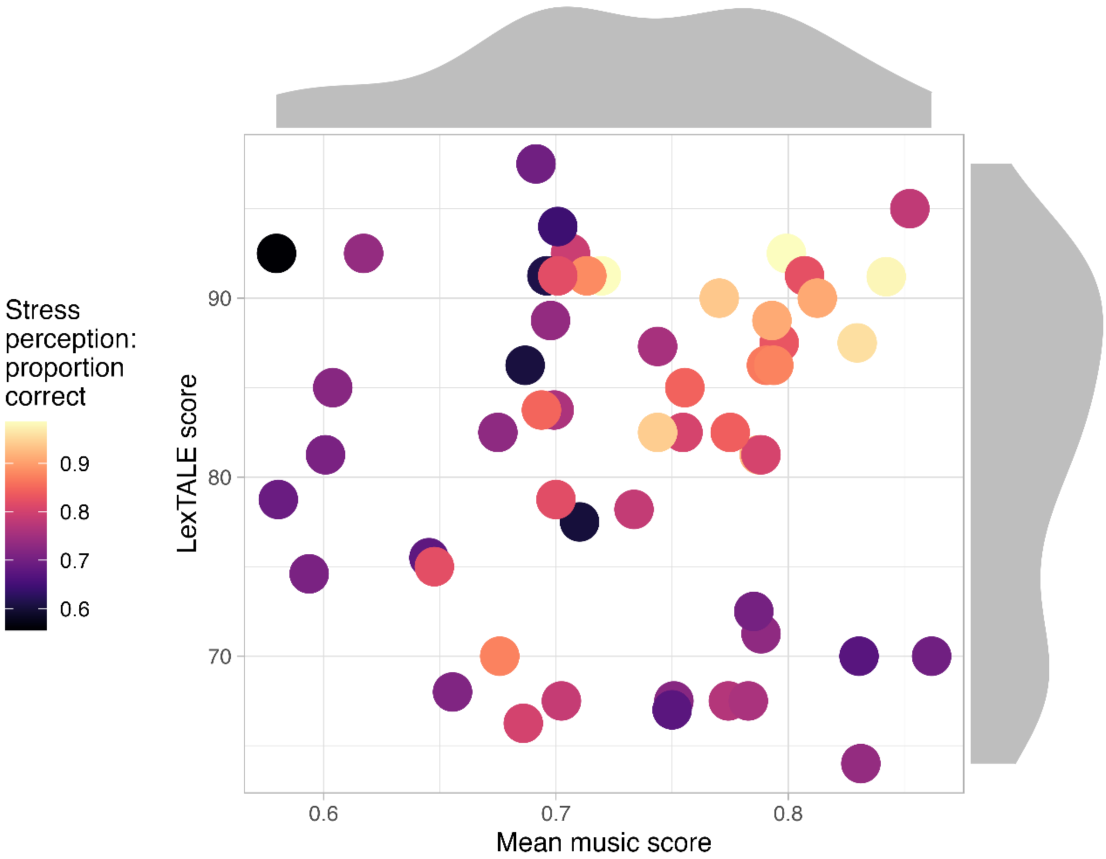


Figure 11a: Scatter plot of mean music scores in proportions (x-axis) vs. LexTALE scores (y-axis). Each dot represents the results of one participant. Colour indicates stress perception accuracy (in proportions), from 0.6 (dark) to 0.9 (light), meaning the darker the worse stress perception score. Margins show the density plots.

The scores visible in the scatter plot in Figure 11a, where each dot represents the overall LexTALE, music and stress results of one participant, are well-distributed across the entire L2-proficiency (LexTALE) by music-proficiency space, indicating a diverse pool of participants with varying levels of English proficiency and musical aptitude. The density plots on the margins further illustrate this variability. Having participants with different combinations of L2 proficiency and musical skills was important in order to be able to address the present research questions. Had the participants been too homogeneous in their L2 proficiency and/or musical perception skills, the potential influence of these dimensions on L2 stress perception would be harder to determine. Notably, figure 11a shows a trend where participants with both higher LexTALE and music perception scores (upper right corner) exhibit lighter colours, signifying better L2 English stress perception accuracy. Conversely, participants farther from the upper right, represented by darker colours, show poorer stress perception scores. Note that neither a high LexTALE score alone, nor a high overall musical perception score alone ensured a relatively high L2 perception success, as several participants in the upper left and lower right quadrants of the space, respectively, had relatively low proportions correct on the stress perception task. This suggests that a combination of high L2 English proficiency and strong musical perception skills is associated with better stress perception, whereas having only one of these proficiencies does not guarantee success in L2 English stress perception.

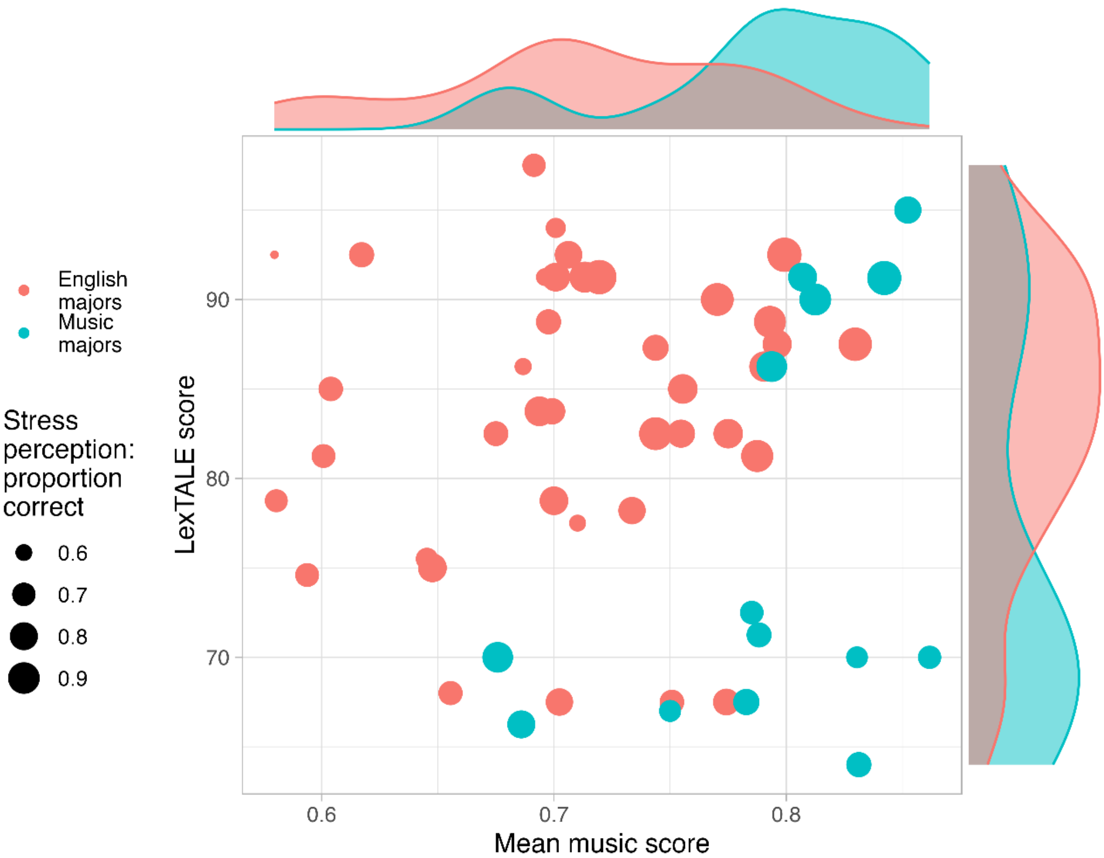


Figure 11b: Scatter plot of mean music scores in proportions (x-axis) vs. LexTALE scores (y-axis). Each dot represents the results of one participant, for English majors (red) and Music majors (green). The size of each dot indicates stress perception accuracy, with larger dot representing higher stress perception accuracy (in proportions). Marginal density plots show the distribution of scores for each group.

Figure 11b is a copy of Figure 11a, adding one extra variable, namely the participants' academic major, English Philology vs Musicology. As expected, English majors generally achieved higher LexTALE scores but lower mean music scores, whereas Music majors show the opposite trend. This distribution indicates that it was fortunate to have recruited participants amongst these two student groups, as this ensured that there are no systematic gaps in the LexTALE scores by music perception scores space.

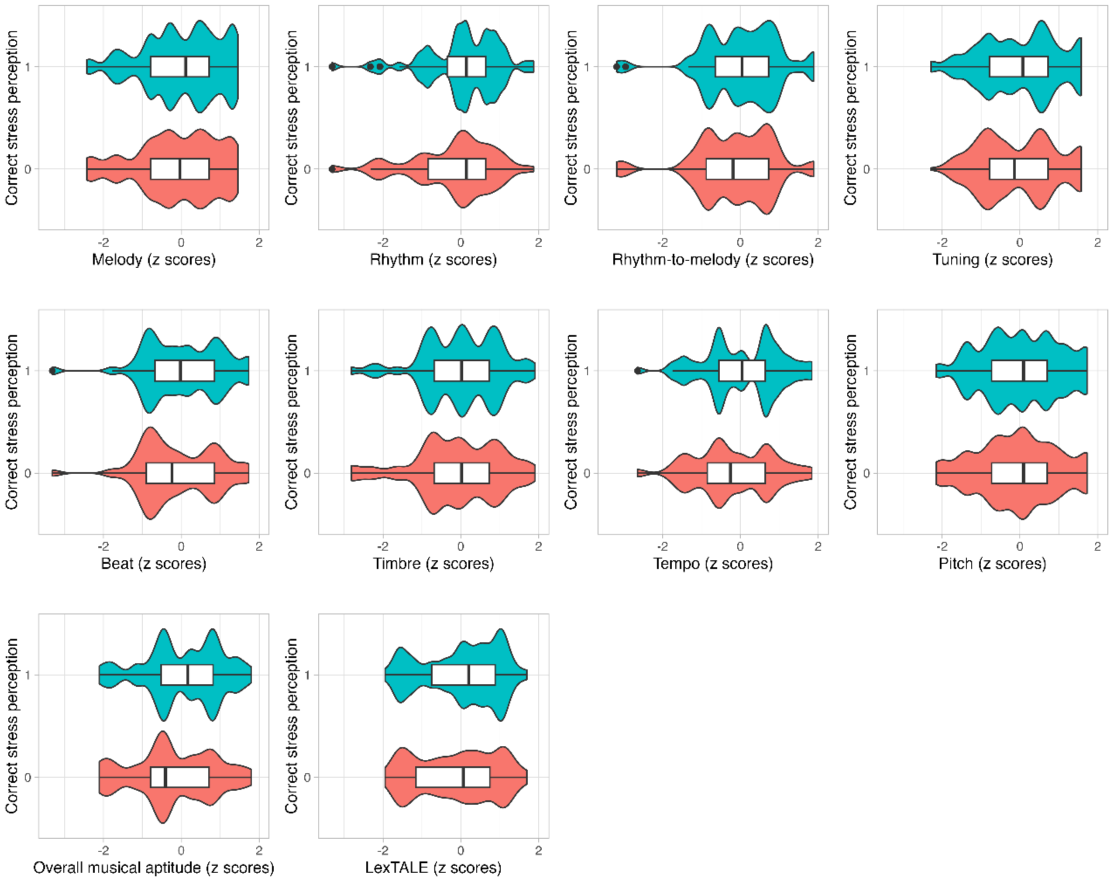


Figure 12: Violin plots with overlaid boxplots showing all the correct (1) versus incorrect (0) stress placement discrimination responses (y-axis) by the participants' scores in the different musical subtests, the overall musical aptitude scores and the LexTALE scores in *z* scores (x-axes).

Figure 12 displays the distributions of the participants' LexTALE scores and of the musical scores for each response elicited in the L2 English stress placement discrimination task. If there is indeed a relationship between LexTALE scores and/or musical perception skills scores on the one hand and the probability of correct perceptual stress placement discrimination on the other hand, then participants with varying LexTALE or musical scores should vary in the probability of producing a correct stress perception response. Consequently, a difference in the frequency distributions of correct versus incorrect stress perception responses (1, in green, vs 0, in red, respectively) within each panel in Figure 12 should be observable, such that the frequency distributions should be more skewed towards the above-average (i.e., positive *z* score) values and away from the below-average (i.e., negative *z* score) values for the correct responses compared to the incorrect responses. Inspecting Figure 12, the shape of the LexTALE violin plots suggests slight differences in the expected direction, namely for the above-average LexTALE scores, the violins are somewhat wider for the correct responses. This trend is also visible in the Overall musical aptitude plot, where the boxplot is shifted to the right for correct responses and the violins are wider for the correct responses for the above-average values. As for the musical subscores, the differences in the frequency distributions are more pronounced for rhythm and tempo, indicating a stronger relationship with correct stress perception responses. The distributions for tuning and beat also show considerable differences between correct and incorrect responses. However, the differences are milder for melody, rhythm-to-melody, timbre, and pitch.

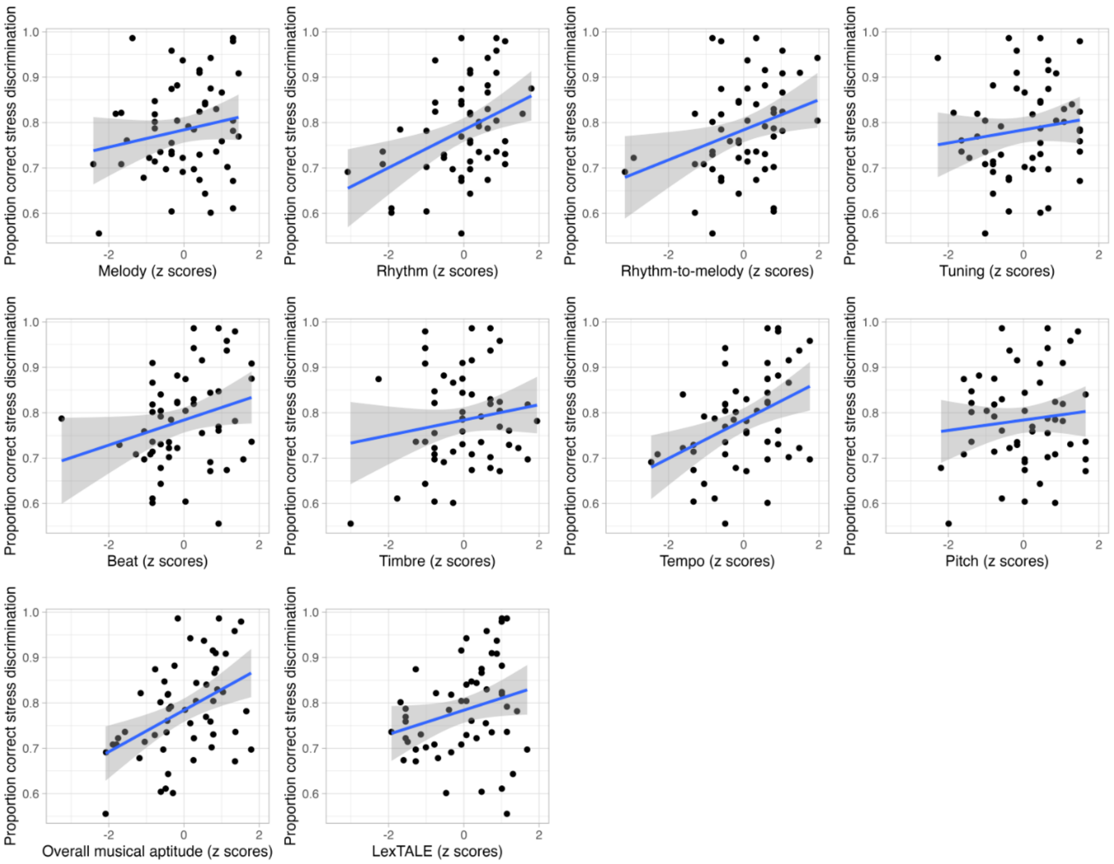


Figure 13: Scatter plots (with simple linear regression lines) showing each participant's mean proportion correct on the stress placement discrimination task (y-axis) by their scores in the different musical subtests, the overall musical aptitude scores and the LexTALE scores in *z* scores (x-axes).

The scatter plots in Figure 13 display each participant's overall proportion of correct responses on the L2 English stress perception task (y axis) and their scores on the LexTALE task, their overall musical perception scores, as well as their different musical subscores. Each plot includes a simple regression line with a shaded confidence interval to highlight a potential relationship between the variables displayed. As expected, a positive correlation seems to exist between stress perception and both the overall music score and the LexTALE score (last two panels in Figure 13). Focusing on the slopes of the regression lines for the specific musical subscores (the first 8 panels in Figure 13), each of the regression lines rises at least a little towards the right, suggesting a positive association between the variables included. The slopes are very mild for melody, tuning, timbre and pitch, whereas they are somewhat steeper for rhythm, rhythm-to-melody, beat and tempo, suggesting a stronger link between these scores and L2 stress perception.

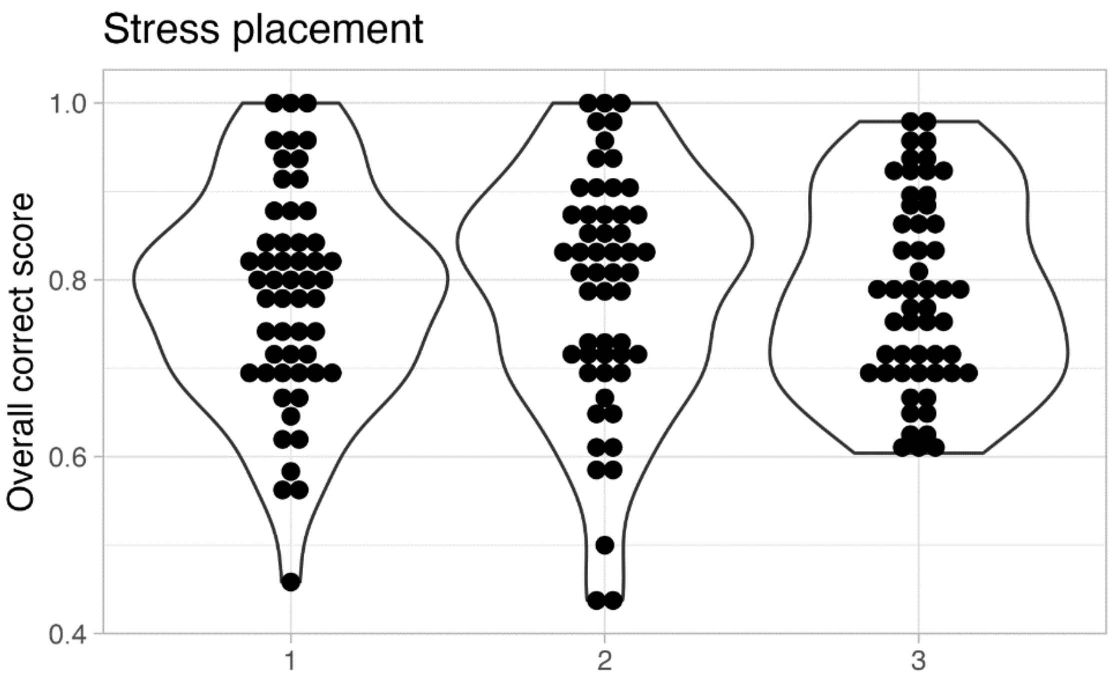


Figure 14: Violin plots and stacked dots plots showing each participant's proportion correct stress discrimination split by the actual placement of stress in the X stimulus (1st, 2nd, or 3rd syllable, on the x axis) within the AXB triplets on the y axis. Each dot represents one participant's proportion correct for each stress placement.

Figure 14 presents the distribution of participants' correct stress discrimination scores based on the placement of stress in the X stimulus (1st, 2nd, or 3rd syllable) within the AXB triplets. The violin plots illustrate the density of the proportion of correct responses, while the stacked dots represent individual participant scores. The y-axis shows the overall correct score, ranging from 0.4 to 1.0, indicating the proportion of correct stress perception. The plot indicates that participants tend to perceive stress more accurately when it is placed on the second syllable compared to the first syllable, which in turn is slightly easier to notice than stress on the third syllable. In contrast, the results of this study differ from those of Tlolková (2018), who found that participants more easily noticed stress when it was on the first syllable compared to the other syllables.

3.2 Mixed-effects logistic regression modelling

First, the response times on the L2 English stress placement AXB discrimination task were inspected, and all responses longer than 4 standard deviations above the average reaction time (approximately 8 seconds) were discarded. This resulted in the removal of 113 responses, leaving 7,493 responses for further analysis. Then the data were submitted to a mixed-effect logistic regression model with 'correct' (0 incorrect, 1 correct) as the response variable. The fixed-effects structure included the participant's LexTALE scores (standardized in *z* scores), their overall musical aptitude scores (standardized in *z* scores), the interaction of the LexTALE and the music scores, 'correctStimXstressPosition', i.e. the correct stress placement in the X stimulus (1st, 2nd, or 3rd syllable, sum-coded), the participant's years of learning English as an L2 (in years), and their field of study (English philology = AF or Musicology = MU, sum-coded). The random effects were Participant and the X stimulus, both modelled with varying intercepts. The model formula used in the R script is given in (1).

(1) simple.mdl <- glmer(correct ~ lexTale_z * music_z + correctStimXstressPosition + yearsOfLearning + fieldOfStudy + (1|participant) + (1|stimX), data = data, family = "binomial", control = glmerControl(optimizer = "bobyqa"))

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	1.571	0.697	2.253	0.024
lexTale_z	0.161	0.087	1.856	0.063
music_z	0.40	0.094	4.247	<0.0001
correctStimXstressPosition1	-0.055	0.102	-0.538	0.591
correctStimXstressPosition2	0.007	0.102	0.069	0.945
yearsOfLearning	-0.009	0.053	-0.174	0.862
fieldOfStudy1	0.144	0.115	1.253	0.21
lexTale_z:music_z	0.298	0.084	3.539	<0.001

Table 12: Coefficients estimated by a mixed-effects logistic regression model.

Table 12 gives the model coefficient estimates. The model confirmed that the overall music score had reliable effect as a predictor on the probability of a correct answer on the stress discrimination task (logit est.: 0.40, $SE = 0.094$, $z = 4.25$, $p = 2.17 \times 10^{-5}$). At the same time LexTALE was not found to predict correct stress discrimination reliably on its own (i.e. from this model, sufficient confidence about its effect is not obtained), however it interacted with the overall music score (logit est.: 0.298, $SE = 0.084$, $z = 3.54$, $p < 0.001$). To visualize the reliable interaction between the LexTALE score and the overall music aptitude score, see Figure 15.

Which syllable was stressed in the X stimulus (1st, 2nd, or 3rd syllable) did not matter reliably for the probability of giving a correct answer. Therefore, the results of the present thesis do not replicate Tlolková's (2018) results which suggested that it was easier for participants to notice stress, when it was on the first as opposed to the other syllables. At the same time, no reliable effect of the years of learning was found in the present study and no evidence in the data that fields of study on their own would matter for producing a correct vs incorrect response on the stress perception task.

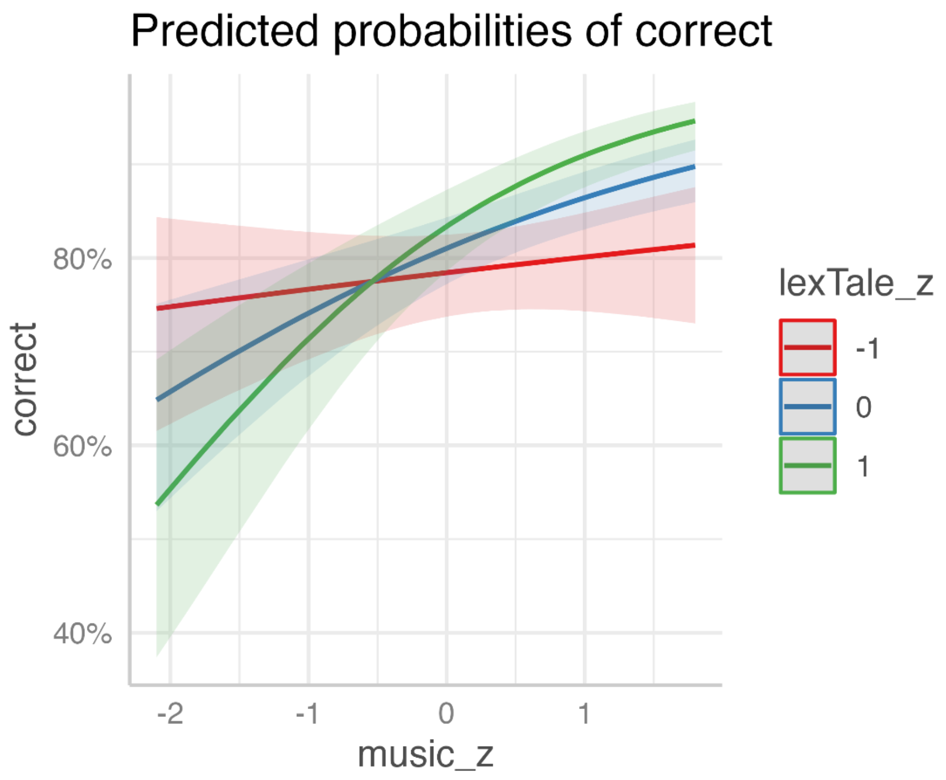


Figure 15: Predicted probabilities of a correct response on the stress discrimination task (y-axis) as a function of the standardized LexTALE scores (colour bands) and overall music scores (x-axis). LexTALE scores are standardized into three levels (-1, 0, 1) and visualized as red, blue, and green bands, respectively. The ribbons represent 95% confidence intervals.

Whereas for the responses produced by participants with a LexTALE score 1 standard deviation below the average (depicted in red), no significant effect of the musical scores on the probability of correct responses is observed. In contrast, for participants with average LexTALE scores (in blue) and those with above-average LexTALE scores (in green), the probability of a correct response rises with rising musical scores. The analysis revealed that overall musical score is a reliable predictor of correct L2 English stress discrimination (also in interaction with L2 proficiency, as indexed by the LexTALE scores).

In order to test statistically which specific musical subscores predicted correct stress discrimination, another mixed-effect logistic regression model was fitted on the same data. Again, the modelled response variable was ‘correct’ (0 incorrect, 1 correct). The fixed effects were the participant’s LexTALE score (standardized in *z* scores), each of the eight music subscores (standardized in *z* scores) and each of their individual interaction with LexTALE. As in the previous model, the random effects were Participant and the X stimulus, both modelled with varying intercepts. The model formula used in the R script is given in (1).

(2) detailed.mdl <- glmer(correct ~ lexTale_z + melody_z + melody_z:lexTale_z + rhythm_z + rhythm_z:lexTale_z + rhythmMelody_z + rhythmMelody_z:lexTale_z + tuning_z + tuning_z:lexTale_z + beat_z + beat_z:lexTale_z + timbre_z + timbre_z:lexTale_z + tempo_z + tempo_z:lexTale_z + pitch_z + pitch_z:lexTale_z + (1|participant) + (1|stimX), data = data, family = “binomial”, control = glmerControl(optimizer = “bobyqa”))

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.507	0.092	16.326	<0.0001
lexTale_z	0.194	0.064	3.024	0.002
melody_z	0.155	0.084	1.85	0.064
rhythm_z	0.199	0.078	2.559	0.011
rhythmMelody_z	0.063	0.078	0.805	0.421
tuning_z	-0.022	0.089	-0.243	0.808
beat_z	0.158	0.076	2.079	0.038
timbre_z	-0.029	0.07	-0.413	0.679
tempo_z	0.237	0.071	3.31	0.001
pitch_z	0.028	0.085	0.334	0.739
lexTale_z:melody_z	0.141	0.097	1.444	0.149
lexTale_z:rhythm_z	0.173	0.087	1.995	0.046
lexTale_z:rhythmMelody_z	-0.038	0.124	-0.307	0.759
lexTale_z:tuning_z	-0.139	0.105	-1.33	0.183
lexTale_z:beat_z	0.093	0.063	1.47	0.141
lexTale_z:timbre_z	0.099	0.071	1.404	0.16
lexTale_z:tempo_z	0.265	0.077	3.447	0.001
lexTale_z:pitch_z	0.073	0.087	0.844	0.399

Table 13: Coefficients estimated by a mixed-effects logistic regression model.

Table 13 gives the second model's coefficient estimates. The model confirmed that the LexTALE score independently predicts correct stress discrimination (logit estimate: 0.194, SE = 0.064, z = 3.02, p = 0.002), despite this not being evident from the previous model. This supports the hypothesis that higher proficiency in L2 improves performance in L2 tasks. The music subscores that significantly influence stress discrimination are rhythm (logit est.: 0.199, SE = 0.078, z = 2.56, p = 0.011), beat (logit est.: 0.158, SE = 0.076, z = 2.08, p = 0.038) and tempo (logit est.: 0.237, SE = 0.071, z = 3.31, p = 0.001). Furthermore, the interaction effects reveal that both rhythm (logit est.: 0.173, SE = 0.087, z = 2, p = 0.046) and tempo (logit est.: 0.265, SE = 0.077, z = 3.45, p = 0.001) interact significantly with LexTALE scores.

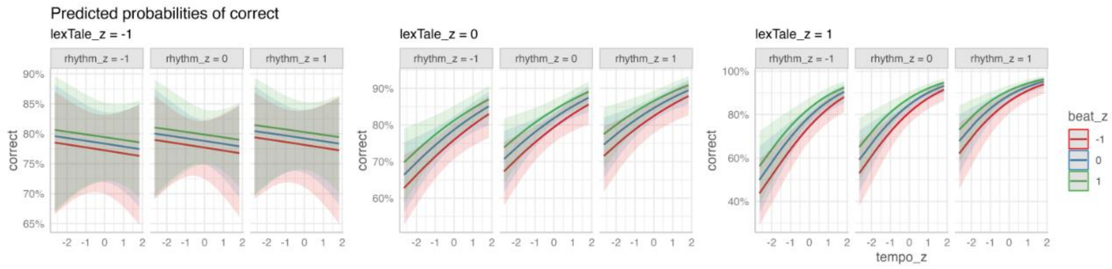


Figure 16a: Predicted probabilities of correct stress discrimination as a function of tempo (*z* scores) across different levels of LexTALE (*z* scores) and rhythm (*z* scores). Each panel represents a different LexTALE score (-1, 0, 1), with lines representing varying levels of rhythm (-1, 0, 1), and colours representing different beat scores (-1: red, 0: blue, 1: green).

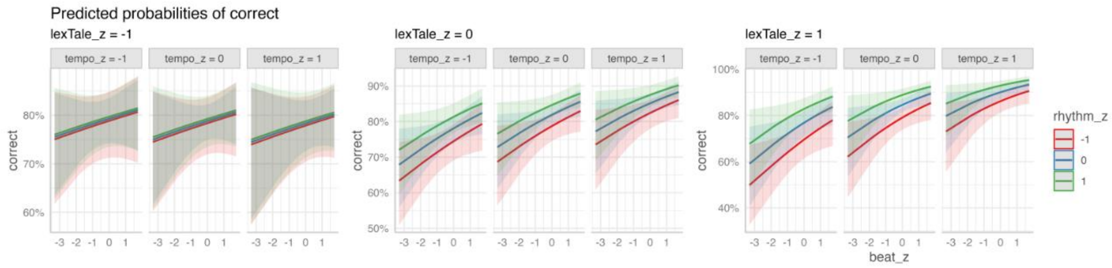


Figure 16b: Predicted probabilities of correct stress discrimination as a function of beat (*z* scores) across different levels of LexTALE (*z* scores) and tempo (*z* scores). Each panel represents a different LexTALE score (-1, 0, 1), with lines representing varying levels of tempo (-1, 0, 1), and colours representing different rhythm scores (-1: red, 0: blue, 1: green).

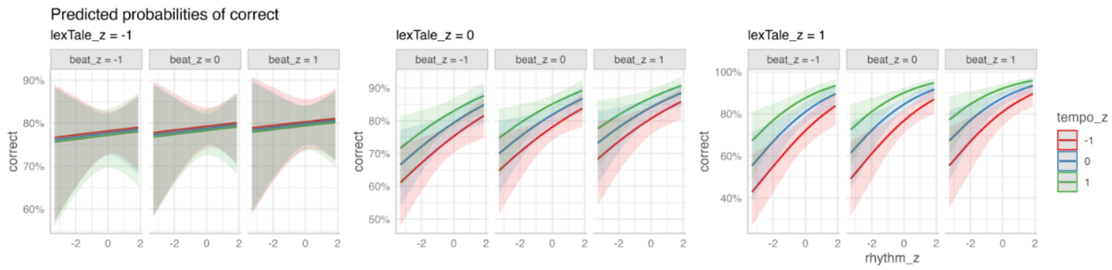


Figure 16c: Predicted probabilities of correct stress discrimination as a function of rhythm (*z* scores) across different levels of LexTALE (*z* scores) and beat (*z* scores). Each panel represents a different LexTALE score (-1, 0, 1), with lines representing varying levels of beat (-1, 0, 1), and colours representing different tempo scores (-1: red, 0: blue, 1: green).

Figures 16a, 16b, and 16c visualize the reliable interactions between the LexTALE scores and the relevant music subscores. Each of these figures show how the probability of correct stress discrimination changes as a function of LexTALE score, the tempo score, the rhythm score, and the beat score. The figures consistently place values modelled for the average LexTALE score in the three centre panels, and those modelled for the LexTALE of 1 standard deviation below-average and above-average in the three left and the three right panels, respectively.

Figure 16a clearly shows the effect of tempo, which is only evident within the average and above-average LexTALE data. The curves in the centre and right panels rise, indicating a positive impact of tempo on stress discrimination accuracy. In contrast, no significant effect of tempo is observed in participants with below-average LexTALE scores. Similarly, figure 16b shows that the effect of beat was moderate in average LexTALE somewhat stronger in above-average LexTALE and weak if any in below-average LexTALE. Finaly, in figure 16c the rhythm effect on stress discrimination is visible in the middle and right panels, where the curves rise, showing a positive correlation between rhythm and correct stress discrimination for participants with average and above-average LexTALE scores. The left panels indicate no significant rhythm effect for participants with below-average LexTALE scores. Taken together, it is observable that the three relevant musical subscores only truly show significant effects in participants with higher proficiency in English, indicating a synergy between the individual facilitating effects of L2 proficiency and the different musical subscores. Overall, these figures reveal that good musical aptitude enhances stress discrimination accuracy primarily in participants with average or above-average English proficiency. Musical aptitude does not significantly aid participants with poor English proficiency. Thus, musical aptitude's positive effects on stress discrimination are contingent upon a certain level of English proficiency, highlighting an interplay between L2 proficiency and musical aptitude in facilitating accurate stress perception.

4 General discussion and conclusion

This thesis supports the connection between musical aptitude and L2 stress perception in Czech learners of L2 English. The experiment comprised three tests: the English stress perception test from Tlolková's thesis (2018), the PROMS-S music aptitude test (Law and Zentner 2012), and the lexical decision task (Lemhöfer & Broersma 2012).

The first hypothesis, that LexTALE scores will be positively associated with L2 stress discrimination, was confirmed. Participants with higher L2 English proficiency exhibited greater success in correctly discriminating L2 English stress. The second hypothesis, which proposed a positive association between PROMS-S scores and L2 stress discrimination, was also supported. Higher overall musical aptitude coincided with greater success in correctly discriminating L2 English stress. This result is consistent with previous research by Gralińska-Brawata and Rybińska (2017), Balčytytė-Kurtinienė (2015), and Garcia and Schwab (2023), indicating a robust link between musical aptitude and linguistic stress perception.

The third hypothesis examined the relationship between specific musical domains and L2 English stress perception. While previous studies focusing on musical domains and phonological awareness identified tonal (melody, pitch) and temporal (rhythm) domains as significant (Culp 2017; Rodríguez 2021; Nardo and Reiterer 2009), this thesis found that only the temporal (rhythm and tempo) and dynamic (beat) domains significantly influenced stress discrimination. The absence of a significant relationship for the tonal (melody, pitch) and spectral/qualitative domains (timbre, tuning) was unexpected but can be attributed to several factors. This is not surprising for spectral domains, as measurements of the AXB discrimination task stimuli showed no significant differences in spectral tilt, F1, or F3 and therefore were not straightforward stress cues. However, it was presumed that tonal domains such as melody and pitch would play a role, given the F0 differences between stressed and unstressed syllables in the stimuli. The pitch variations in English stress patterns might not be as pronounced or easily detectable by non-native speakers, especially if their native language does not use pitch in a similar way. There might have also been a ceiling effect in pitch discrimination, where participants already had high proficiency in pitch sensitivity, resulting in variations that did not significantly impact stress perception. Furthermore, participants might have had more training or exposure to rhythm and tempo through music education or natural language use, making these cues more salient. Processing multiple acoustic cues simultaneously can be cognitively demanding, leading participants to prioritize cues that are less complex to process, such as rhythm and tempo, over melody and pitch. Lastly, variability in individual musical training and cognitive abilities might lead to differences in how musical domains influence language perception, suggesting a need for personalized approaches in future studies.

Participants exhibited a wide range of English proficiency and musical aptitude, contributing to the diversity of the data. This variety was crucial for addressing the research questions, as it provided a broad spectrum of abilities to analyse. Notably, participants with both high LexTALE and music perception scores demonstrated better L2 stress perception accuracy, indicating that a combination of these skills is beneficial. However, proficiency in only one area did not guarantee success in stress perception tasks. Further analysis distinguished between English majors and Music majors, revealing expected trends: English majors had higher LexTALE scores but lower music scores, while Music majors showed the opposite pattern. This balanced distribution was advantageous for the study, as it ensured a comprehensive analysis of the interplay between L2 proficiency and musical skills.

The mixed-effects logistic regression analysis confirmed that overall musical aptitude is a reliable predictor of correct stress discrimination, especially when interacting with L2 proficiency. While LexTALE scores alone did not consistently predict correct responses, their interaction with musical scores did. This highlights the importance of considering both linguistic and musical abilities in understanding L2 stress perception. This finding is counterintuitive, as one might assume that a good musical ear (higher musical aptitude) helps learners at the beginning stages of learning L2 English. However, the present thesis finds that this is not necessarily the case. Even when a student of L2 English has good musical aptitude, it does not assist them in learning the language if they do not possess sufficient language knowledge. Therefore, to effectively leverage better musical aptitude, learners must have a minimal level of proficiency in the language.

The findings of this thesis have several important implications. Firstly, they suggest that musical aptitude can enhance L2 stress perception, but this effect is contingent on a minimum level of L2 proficiency. This implies that integrating musical training in language education might be particularly beneficial for learners who have already attained a basic proficiency level in the target language. Educators could leverage musical exercises to reinforce language skills, potentially improving stress perception and overall pronunciation. Moreover, the differential impact of various musical domains on stress perception highlights the need for targeted musical training. Since rhythm and temporal skills are particularly influential, language educators might consider incorporating rhythmic exercises and temporal pattern recognition activities into their curriculum.

Despite the significant findings, this study has several limitations. The sample size was relatively small, which may affect the generalizability of the results. Future research should aim to replicate these findings with a larger and more diverse sample to confirm the robustness of the observed relationships. Additionally, the study focused solely on Czech learners of L2 English. It remains unclear whether these findings can be generalized to learners from different linguistic backgrounds. Cross-linguistic studies are needed to determine whether the observed relationships hold true for learners of other L2s. Future research could explore the impact of specific musical training interventions on L2 stress perception. Longitudinal studies tracking learners over time would provide valuable insights into how musical aptitude and language proficiency interact during the language learning process. Further investigation into the role of individual differences in musical perception skills is also warranted. Understanding how different musical domains contribute to language learning can inform the design of more effective, personalized language instruction programs.

In conclusion, while musical aptitude positively influences L2 stress perception, its effectiveness is moderated by the learner’s language proficiency. This nuanced understanding challenges the simplistic notion that musical skills uniformly benefit language learning and highlights the complex interplay between different cognitive domains.

5 Works cited

Balčytytė-Kurtinienė, Giedrė. 2015. "Musical aptitude and language: activating phonetic skills through music." 8th Conference on ICT in Language Learning: 417-421.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67, no. 1: 1-48. https://doi.org/10.18637/jss.v067.i01.

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. 2003. "Effects of Disfluencies, Predictability, And Utterance Position on Word Form Variation in English Conversation." Online. *The Journal of The Acoustical Society of America* 113 (2): 1001-1024. https://doi.org/10.1121/1.1534836.

Bentley, Andrew. 1966. *Musical Ability in Children and Its Measurement*. London, England: Harrap.

Boersma, Paul and David Weenink. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.3.03, available at http://www.praat.org/

Carroll, John B. 1993. *Human Cognitive Abilities. A Survey of Factor-Analytic Studies*. Cambridge University Press.

Chobert, J., C. François, J.-L. Velay, and M. Besson. 2014. "Twelve Months of Active Musical Training In 8- To 10-Year-Old Children Enhances the Preattentive Processing of Syllabic Duration and Voice Onset Time." *Cerebral Cortex* 24 (4): 956-967. doi:10.1093/cercor/bhs377.

Christiner, Markus, and Susanne Reiterer. 2018. "Early Influence of Musical Abilities and Working Memory On Speech Imitation Abilities: Study with Pre-School Children". Online. *Brain Sciences* 8 (9): 1-16. doi:10.3390/brainsci8090169.

Culp, Mara E. 2017 "The Relationship Between Phonological Awareness and Music Aptitude." *Journal of Research in Music Education* 65 (3): 328–46. https://www.jstor.org/stable/48588624.

Duběda, Tomáš, and Jan Votrubec. 2005. "Acoustic Analysis of Czech Stress: Intonation, Duration and Intensity Revisited." Online. *Interspeech 2005*, 1429-1432. https://doi.org/10.21437/Interspeech.2005-509.

Eriksson, Anders, and Mattias Heldner. 2015. "The Acoustics of Word Stress in English as A Function of Stress Level and Speaking Style." Online. *Interspeech 2015*, 41-45. https://doi.org/10.21437/Interspeech.2015-9.

Fonseca-Mora, Carmen, Carmen Toscano-Fuentes, and Kathleen Wermke. 2011. "Melodies that help: The relation between language aptitude and musical intelligence." International Journal of English Studies 22.1: 101-118.

François, C., J. Chobert, M. Besson, and D. Schon. 2013. "Music Training for The Development of Speech Segmentation." *Cerebral Cortex 23* (9): 2038-2043. doi:10.1093/cercor/bhs180.

Garcia M.T. M. and Sandra Schwab. 2023. "Relation between musical aptitude and L2 stress perception in French- and Korean- speaking listeners." In: Radek Skarnitzl & Jan Volín (Eds.), Proceedings of the 20th International Congress of Phonetic Sciences: 2378-2382. Guarant International.

Gralińska-Brawata, A. and Paulina Rybińska. 2017. “The Relationship between the Production of Word Stress and Musical Abilities in Polish Learners of English.” *Research in Language* 15 (3): 265-283. doi:10.1515/rela-2017-0015.

Gordon, Edwin E. 1979. *Primary Measures of Music Audiation*. Chicago: G. I. A. Publications.

Gordon, Edwin E. 1982. *Intermediate Measures of Music Audiation.* Chicago: G. I. A. Publications.

Gordon, Edwin E. 1989a. *Advanced Measures of Music Audiation.* Chicago: G. I. A. Publications.

Gordon, Edwin E. 1989b. *Audie: A Game for Understanding and Analyzing Your Child's Music Potential.* Chicago: GIA Publications, Inc. (Pad of game sheets included).

Gordon, Edwin E. 1995. *Manual: Musical Aptitude Profile*. 3rd ed. Chicago, IL: GIA Publications.

Gordon, Edwin E. 2007. *Learning Sequences in Music: A Contemporary Music Learning Theory*. 2007 Edition. Chicago: GIA Publications.

Hyde, Krista L., Jason Lerch, Andrea Norton, Marie Forgeard, Ellen Winner, Alan C. Evans, and Gottfried Schlaug. 2009. “Musical Training Shapes Structural Brain Development.” *The Journal of Neuroscience* 29 (10): 3019-3025. doi:10.1523/JNEUROSCI.5118-08.2009.

Jekiel, Mateusz and Kamil Malarski. 2021. “Musical Hearing and Musical Experience in Second Language English Vowel Acquisition.” *Journal of Speech, Language and Hearing Research (Online)* 64 (5): 1666-1682. https://doi.org/10.1044/2021_JSLHR-19-00253.

Kang, Robert, Grace Liu Nimmons, Ward Drennan, Jeff Longnion, Chad Ruffin, Kaibao Nie, Jong Ho Won, Tina Worman, Bevan Yueh, Jay Rubinstein. 2009. “Development and Validation of the University of Washington Clinical Assessment of Music Perception Test.” *Ear Hear* 30: 411–418.

Klatt, D. H. 1976. “Linguistic uses of segmental duration in English: Acoustic and perceptual evidence.” *The Journal of the Acoustical Society of America,* 59(5): 1208-1221.

Ladefoged, Peter. 1996. *Elements of Acoustic Phonetics*. 2nd ed. The University of Chicago Press.

Ladefoged, Peter, and Keith Johnson. 2011. *A Course in Phonetics*. Sixth Edition. Boston: Wadsworth/Cengage Learning.

Law, Lily N. C., and Marcel Zentner. 2012. “Assessing Musical Abilities Objectively: Construction and Validation of The Profile of Music Perception Skills.” Online. *Plos One* 7 (12). https://doi.org/10.1371/journal.pone.0052508.

Lehiste, Ilse, and Robert Allen Fox. 1992. “Perception of Prominence by Estonian and English Listeners.” *Language and Speech*. journals.sagepub.com/doi/abs/10.1177/002383099203500403.

Lemhöfer, Kristin, and Mirjam Broersma. 2012. “Introducing LexTALE: A Quick and Valid Lexical Test for Advanced Learners of English.” Online. In *Behavior Research Methods*, 44:325-343. https://doi.org/10.3758/s13428-011-0146-0.

Llanes-Coromina, Judith, Pilar Prieto, and Patrick Rohrer. 2018. “Brief Training with Rhythmic Beat Gestures Helps L2 Pronunciation in A Reading Aloud Task.” *Speech Prosody* 2018. ISCA: ISCA, 498-502. doi:10.21437/SpeechProsody.2018-101.

Lüdecke, Daniel. 2018. "ggeffects: Tidy Data Frames of Marginal Effects from Regression Models." *Journal of Open Source Software* 3, no. 26: 772. https://doi.org/10.21105/joss.00772.

Marques, Carlos, Sylvain Moreno, São Luís Castro, and Mireille Besson. 2007. "Musicians Detect Pitch Violation in A Foreign Language Better Than Nonmusicians: Behavioral and Electrophysiological Evidence." *Journal Of Cognitive Neuroscience* 19 (9): 1453-1463. doi:10.1162/jocn.2007.19.9.1453.

Machač, Pavel, and Radek Skarnitzl. 2009. *Principles of Phonetic Segmentation*. Praha: Epocha.

Milovanov, Riia, Minna Huotilainen, Vesa Välimäki, Paulo A.A. Esquef, and Mari Tervaniemi. 2008. "Musical Aptitude and Second Language Pronunciation Skills in School-Aged Children: Neural and Behavioral Evidence." *Brain Research* 1194: 81-89. doi:10.1016/j.brainres.2007.11.042.

Milovanov, Riia, Päivi Pietilä, Mari Tervaniemi, and Paulo A.A. Esquef. 2010. "Foreign Language Pronunciation Skills and Musical Aptitude: A Study of Finnish Adults with Higher Education." *Learning and Individual Differences* 20 (1): 56-60. doi:10.1016/j.lindif.2009.11.003.

Müllensiefen, Daniel, Bruno Gingras, Jason Musil, and Lauren Stewart. 2012. "A New Tool for Measuring Musical Sophistication: The Goldsmiths Musical Sophistication Index." Presentation given at the 12th International Conference on Music Perception and Cognition (ICMPC), July 25, Thessaloniki.

Nardo, Davide, and Susanne Maria Reiterer. 2009. "Musicality and Phonetic Language Aptitude." In *Language Talent and Brain Activity*, edited by Grzegorz Dogil and Susanne Maria Reiterer, 213-256. Berlin: De Gruyter Mouton. doi:10.1515/9783110215496.213.

Ortega-Llebaria, Marta, and Pilar Prieto. 2011. "Acoustic Correlates of Stress in Central Catalan and Castilian Spanish." Online. *Language And Speech* 54 (1): 73-97. https://doi.org/10.1177/0023830910388014.

Palková, Zdena. 1994. *Fonetika a fonologie češtiny s obecným úvodem do problematiky oboru*. Prague: Karolinum

Pastuszek-Lipińska, Barbara. 2008. "Musicians Outperform Nonmusicians in Speech Imitation." *Lecture Notes in Computer Science* 4969: 56–73. https://doi.org/10.1007/978-3-540-85035-9_4.

Peretz, Isabelle, Anne-Sylvie Champod, and Krista L. Hyde. 2003. "Varieties of Musical Disorders: The Montreal Battery of Evaluation of Amusia." *Annals of the New York Academy of Sciences* 999: 58–75.

Perfors, Amy, and Jia Hoong Ong. 2012. "Musicians Are Better at Learning Non-Native Sound Contrasts Even in Non-Tonal Languages". Online. *Proceedings Of the Annual Meeting of The Cognitive Science Society* 34: 839-844. https://escholarship.org/uc/item/9g9917w5.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Roach, Peter. 1998. *English Phonetics and Phonology*. Cambridge: Cambridge University Press.

Robert Slevc, L. and A. Miyake. 2006. "Individual Differences in Second-Language Proficiency: Does Musical Ability Matter?" *Psychological Science* 17 (8): 675-681. doi:10.1111/j.1467-9280.2006.01765.x.

Rodríguez, Andrés Pujazón. 2021. "Musical Aptitude and foreign language receptive pronunciation." *Phonica* 17: 72-89.

Rossi, Sonja, Manfred F. Gugler, Markus Rungger, Oliver Galvan, Patrick G. Zorowka, and Josef Seebacher. 2020. "How the Brain Understands Spoken and Sung Sentences." *Brain Sciences* 10 (1): 36. https://doi.org/10.3390/brainsci10010036.

Seashore, Carl, Donald Lewis, and Joseph G. Saetveit. 1960. *A Second Revision of the Manual of Instructions and Interpretations for the Seashore Measures of Musical Talents* (1939 revision). New York, NY: The Psychological Corporation.

Skarnitzl, Radek, and Anders Eriksson. 2017. "The Acoustics of Word Stress in Czech as A Function of Speaking Style." Online. *Interspeech 2017*, 3221-3225. https://doi.org/10.21437/Interspeech.2017-417.

Skarnitzl, Radek, and Jana Rumlová. 2019. "Phonetic Aspects of Strongly-Accented Czech Speakers of English." *Auc Philologica* 2019 (2): 109-128. doi:10.14712/24646830.2019.21.

Talamini, Francesca, Massimo Grassi, Enrico Toffalini, Rosa Santoni, and Barbara Carretti. 2018. "Learning a Second Language: Can Music Aptitude or Music Training Have a Role?" Learning and Individual Differences 64: 1-7. doi:10.1016/j.lindif.2018.04.003.

Tlolková, Hana. 2018. "Stress 'Deafness' of Czech Learners of English." Master thesis, Palacký University Olomouc.

Volín, Jan, and Lenka Weingartová. 2014. "Acoustic Correlates of Word Stress as A Cue to Accent Strength." Online. *Research in Language* 12 (2): 175-183. https://doi.org/10.2478/rela-2014-0008.

Wallentin, Mikkel, Andreas Højlund Nielsen, Morten Friis-Olivarius, Christian Vuust, and Peter Vuust. 2010. "The Musical Ear Test, a New Reliable Test for Measuring Musical Competence." *Learning and Individual Differences* 20: 188–196.

Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.

Wing, Henry. 1948. *Tests of Musical Ability and Appreciation: An Investigation into the Measurement, Distribution, and Development of Musical Capacity*. London, England: Cambridge University Press.

Wong, Patrick C. M., Erika Skoe, Nicole M. Russo, Tasha Dees, and Nina Kraus. 2007. "Musical Experience Shapes Human Brainstem Encoding of Linguistic Pitch Patterns". *Nature Neuroscience* 10 (4): 420-422. doi:10.1038/nn1872.

Zybert, Jerzy, and Sabina Stępień. 2009. "Musical Intelligence and Foreign Language Learning." *Research in Language* 7 (December): 99-111. doi:10.2478/v10015-009-0007-4.

6 Appendices

In the section bellow you can find the data it was referred to throughout the thesis. The data are organised as follows.

Appendix A: Informed Consent

This thesis is accompanied by an electronic appendix titled “Results, R script” and “Stress discrimination task – measured stimuli”. The first folder includes the results of all three tests, along with the figures and the R script. The second folder titled contains the measured stimuli from the AXB stress discrimination test, as well as the corresponding figures and tables.

Appendix A: Informed Consent

Faculty of Arts

Informovaný souhlas k účasti ve výzkumu diplomové práce

Autor práce: Helena Rychetská (helena.rychetska01@upol.cz)

Vedoucí práce: Václav Jonáš Podlipský (vaclav.j.podlipsky@upol.cz)

Děkuji vám za váš zájem o účast v mém výzkumu. Předtím, než začnete, si prosím pečlivě přečtěte následující souhlas:

Byl(a) jsem informován(a) o výzkumu diplomové práce na Katedře anglistiky a amerikanistiky a měl(a) jsem možnost pokládat otázky a získat potřebné informace. Jsem si vědom(a) toho, že účast na výzkumu zahrnuje poslech různých zvuků a jejich srovnávání, krátký dotazník (v rámci hudebního testu) a poslech neexistujících anglických slov, rozlišování důrazu ve slabikách a krátký lexikální test (v rámci anglického testu). Výzkum probíhá ve dvou částech ve dvou různých dnech – hudební část trvá přibližně 45 minut a anglická asi 30 minut. Účast na testu nezahrnuje žádná rizika nebo negativní dopady na zdraví účastníka. Jsem si vědom(a) toho, že data budou anonymizována a využita pouze pro vědecké účely. Má účast je dobrovolná a mohu svůj souhlas kdykoliv odvolat. Pokud se během experimentu rozhodnu, že si nepřeji pokračovat, lze kdykoliv přerušit účast na výzkumu bez udání důvodu. Rozumím, že mohu pro další informace kontaktovat autora nebo vedoucího práce.

Stvrzuji, že jsem byl(a) informován(a) o průběhu experimentu, jeho účelu, anonymitě a dobrovolné povaze na jeho účasti. Beru na vědomí, že neexistují žádná rizika spojená s účastí v tomto experimentu.

Souhlasím se všemi uvedenými informacemi a s podmínkami účasti:

Podpis: ____ Datum: ___________

Palacký University Olomouc | Faculty of Arts
Křížkovského 511/10 | CZ 771 47 Olomouc | Czech Republic | T: +420 585 633 620
www.ff.upol.cz