Filozofická fakulta Univerzity Palackého

Does Non-Tonal L1 Background Affect Musical Pitch Skills?

Musical Abilities of Czech and English Native Speakers

(Bakalářská práce)

2024

Vojtěch Kopecký

Filozofická fakulta Univerzity Palackého

Katedra anglistiky a amerikanistiky

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a uvedl jsem plný seznam citované a použité literatury.

V Olomouci dne 8. 5. 2024 Vojtěch Kopecký

Acknowledgements

I want to express my deepest gratitude to my supervisor Mgr. Václav Jonáš Podlipský, Ph.D. for his guidance, comments, discussions, and support throughout the process of writing this thesis.

Additionally, I want to thank KAAmarádi, Mgr. Markéta Dančová, and Mgr. Daniel Kopecký for mental support and encouragement in times of despair.

Abstract

This thesis reviews previous literature that is concerned with the effect that the native language background has on one's musical abilities. It discusses a possible difference in the language-to-music transfer between Czech and English, two non-tonal languages that differ notably in the width of the F0 contours they employ in everyday speech. An appropriate methodology for an experiment targeting this possible difference is proposed, and future research in this field is encouraged.

Keywords:

Musical skills, pitch perception, pitch production, L1 background, non-tonal languages, speech, music, Czech, English, F0 contours

Table of Contents

Introduction

Previous research has established a strong connection between music and speech. It has been shown that musicianship, musical experience, and musical aptitude provide a significant advantage in second language acquisition (see Slevc and Miyake 2006; Prefors and Ong 2012; Talamini et al. 2018; Wiener and Bradley 2023; Jekiel and Malarski 2021). Conversely, a number of studies have examined the link between music and speech in the opposite direction, i.e., they studied the influence the native language (L1) background has on one's musical abilities (see Pfordresher and Brown 2009; Bidelman et al. 2013; Choi 2021; Chen et al. 2016). A clear connection has been established between having a tone language as one's L1, i.e., a language in which a pattern of the pitch of the voice carries lexical meaning, and musical abilities, especially musical pitch perception (see Pfordresher and Brown 2009; Bidelman et al. 2013; Choi 2021; Chen et al. 2016). According to these studies, tone-language (e.g., Mandarin) background enhances the speaker's musical pitch perception abilities as compared to a speaker of a non-tonal language (e.g., English).

However, a majority of the aforementioned studies focus on the effect of a native tone language on musical pitch perception. Only an insufficient number of studies explore, besides musical pitch perception, also the potential L1 background effects on musical pitch production (see Pfordresher and Brown 2009). In addition, to the best of my knowledge, no study so far has yet tested whether the L1 background modulates both musical pitch perception and production even for native speakers of non-tonal languages, and if so, then how the potential effects vary among native speakers of different non-tonal languages. This is important because non-tonal languages differ in the variability and dynamics of the fundamental frequency (F0) contours they employ. English and Czech differ exactly in this respect. On the whole, Czech has considerably flatter F0 contours than English (Volín et al. 2015; Volín et al. 2017). I have therefore decided to attempt to fill this gap by proposing a research study testing Czech and English speakers' musical pitch perception and production. Firstly, I will provide a detailed overview of existing relevant literature, based on which I will then propose appropriate methodology for an adequate experiment targeting this issue.

As I have already mentioned at the beginning, previous studies, mainly focusing on pitch perception, have provided evidence that tonal language background enhances speakers' musical pitch perception (Pfordresher and Brown 2009; Bidelman et al. 2013; Choi 2021; Chen et al. 2016). Such enhancement may be explained by the speakers' extra need to focus on tone, i.e., the changing pitch contours, at practically any given moment in discourse to retrieve lexical content (Liu et al. 2023, 1916). This is the case both when producing speech, as well as when listening to the speech of others.

A recent mass-sample study (Liu et al. 2023) offers new insights into this topic. The authors performed a meta-analysis of 20 previous studies and conducted their own experiment in which they examined nearly half a million native speakers of tonal, pitch-accented, and non-tonal languages. The results of the meta-analysis were not conclusive, juxtaposing the mixed findings of previous literature supporting (see Liu et al. 2023; Choi 2021; Chen et al. 2016; Ngo et al. 2016; Swaminathan et al. 2021; Bidelman et al. 2013; Zhang et al. 2020; Wong et al. 2012) or not supporting (see Peretz et al. 2011; Stevens et al. 2013; Bidelman et al. 2011a; Chang et al. 2016) the effect of tonal language background on musical abilities. The findings of the mass-sample study support the connection between tonal language background and musical pitch processing. More importantly, though, they differentiated between two main types of pitch processing in a similar manner as only some other previous studies did (see Zatorre and Baum 2012; Bidelman et al. 2011a).

The first type of pitch processing defined by Zatorre and Baum (2012) is the so-called “coarse-grained” perception, which is engaged both in speech and in music, and it focuses on pitch contours, i.e., the patterns of changes of pitch over time (Zatorre and Baum 2012; see also Liu et al. 2023). The second type is the so-called “fine-grained” or “fine-scale” pitch perception, which is most likely unique to music and focuses on smaller pitch differences, such as the musical relationships between individual intervals used in scales (Zatorre and Baum 2012; Liu et al. 2023; Bidelman et al. 2011a). This differentiation of pitch processing suggests that tone language speakers are better at musical pitch processing than non-tonal language speakers only when tested on the “coarse-grained” perception. Additionally, allowing for the proposition that “fine-grained” musical perception is only activated in music, it naturally follows that this finer perception can only be polished through practising one’s musical skills. This reasoning is supported by the results of the mass-sample study by Liu et al. (2023) who showed that tonal language speakers have an advantage over non-tonal language speakers in tasks targeting “coarse-grained” perception; however, they show no advantage in tasks targeting “fine-grained” perception which proves to be enhanced by musical experience and training instead (Liu et al. 2023, 1918-1919).

English and Czech, apart from being non-tonal languages, are also considered intonational languages, i.e., they use intonation on the level of a phrase or a sentence to, e.g., convey non-lexical information such as emphasis, differentiate between questions and statements or regular and infant-directed speech, or to express emotions (Liu et al. 2023, 1917). Volín et al. (2015) argue for the importance of native-like intonation when discussing its role in the comprehensibility of non-native speech. As they note, "…intonation along with other prosodic aspects plays a key role in the improvement of intelligibility" (Volín et al. 2015, 108). However, even though English and Czech speakers do focus on the changes of pitch in running speech to a significant extent, they do not need to give it as close attention as speakers of tonal languages do. That is because tonal language speakers not only need to focus on pitch contours only to convey additional non-linguistic meaning, as is the case with intonational languages, but also to retrieve the intended lexical meaning (Liu et al. 2023, 1917). Native speakers of tone languages are, therefore, honing their "coarse-grained" perception more intensively than speakers of non-tonal languages, leading to the aforementioned advantage. Nevertheless, as already said, English and Czech speakers still do need to pay attention to pitch contours to a significant extent, leaving them some room for enhancing their perception as well.

Given the documented effect of tonal languages on musical pitch skills, I predict that native English speakers have an advantage in musical pitch perception and possibly also production, as compared to native speakers of Czech. Not because English speakers would also have the extra need to focus on pitch during their day-to-day use of their native language to reflect differences between individual lexical morphemes, because that is not the case, but because English is a noticeably more melodic language than Czech, as its F0 contours are much wider than those of Czech (Volín et al. 2015; Volín et al. 2017; Skarnitzl and Hledíková 2022). English F0 contours being more dynamic may potentially lead to greater perceptual sensitivity to pitch in general, i.e., to the F0 of non-speech periodic sounds, including musical tones and harmonies, as well as to the F0 of speech, and possibly also to better control of the vocal folds to produce a desired F0. Based on this reasoning, it is possible to hypothesise that English native speakers are more proficient at musical pitch perception and production than Czech native speakers.

1 Literature Review

1.1 Pitch and F0 and Their Role in Language and Music

This paper investigates how one's linguistic background affects one's musical abilities. It is therefore important to understand melody both in the domain of speech and in the domain of music in greater detail. In this chapter, I will discuss these domains together with the role that pitch, and fundamental frequency (F0) have in each of them. It is apparent that language and music share many features. As Zatorre and Baum (2012) broadly say, both speech and music use the same processes and mechanisms for pitch perception and production, "[b]oth share the features of hierarchical structure, complex and sensorimotor demands, and both are used to convey and influence emotions, among other functions. Both […] also prominently use acoustical frequency modulations, perceived as variations in pitch, as part of their communicative repertoire" (p. 1). Despite the many shared features, speech and music also differ greatly in terms of both pitch perception and production.

1.1.1 Pitch and Fundamental Frequency in Language

Firstly, it is important to note that we differentiate between pitch and fundamental frequency (F0). According to Ladefoged and Johnson (2014), "[t]he pitch of a sound is an auditory property that enables a listener to place it on a scale going from low to high, without considering its acoustic properties" (p. 25), i.e., it is how a particular sound is perceived by an individual listener. Pitch is therefore subjective because sounds that some consider to be "high" on a perceptual scale may not be considered as such by others. F0, on the other hand, is an actual physical property of a sound, specifically, it is the lowest frequency in a periodic signal and the common denominator of the harmonically related higher frequency components. This physical quantity is independent of whether or not the sound is being heard and will therefore be the same even when the sound is heard by multiple different listeners.

In speech, pitch perception and production are the processes of decoding and encoding of the changes in fundamental frequency perceived as pitch contours. Pitch in speech is a part of a larger set of speech parameters called prosody or suprasegmental features which, apart from intonation, also include stress, and rhythm (Zatorre and Baum 2012). As Zatorre and Baum (2012) nicely summarise, the usage of prosody includes "…distinguishing word meanings in tone languages (e.g., Mandarin and Thai), disambiguating sentence structures (e.g., distinguishing questions from statements), highlighting or emphasizing elements in a sentence, and signaling emotion (including irony and sarcasm)" (p. 1). Additionally, some languages also use stress to distinguish between lexical items (e.g., English differentiating between nouns and verbs such as "record"). Skarnitzl and Hledíková (2022) describe that "[p]rosodic features serve the role of organizing the flow of speech and giving it structure by dividing it into smaller units, which are called prosodic phrases, prosodic units, tone units, thought groups etc." (p. 3). Shattuck-Hufnagel and Turk (1996) further define the term "prosodic" or "intonational" phrase as "…the domain of a perceptually coherent intonational contour, or tune" (p. 210) which, at least in English, is "delimited by prosodic boundaries […] which are usually signaled by melodic and temporal features (melodic movements and final deceleration or lengthening, respectively), sometimes by a pause" (Skarnitzl and Hledíková 2022, 3). Finally, since it seems to be prosody that allows us to remember a sequence of words, the degree of clearness of phrasal prosody is an important determiner for the level of ease and accuracy of listeners' speech comprehension, i.e., the clearer the prosodic phrasing, the easier and better the comprehension (Skarnitzl and Hledíková 2022, 4).

An important aspect of the perception and production of prosody is relativity. As Honorof and Whalen (2005, 2193) note, in linguistics, fundamental frequency's absolute value is not of interest. It is its relation to the speaker's range that is important. In other words, speakers are not required to use specific frequencies in their speech. In contrast, in music, it is common to attune one's production to a specific frequency in absolute terms as much as possible. In speech, speakers only need to focus on the rough directions from one F0 frequency to another, i.e. on intonation or pitch contours, in order to convey the intended meaning, be it lexical, grammatical, or extralinguistic (Liu et al. 2023, 1917). As a result, fundamental frequencies and their combinations used during speech are unique to each speaker and to each speech act, although they are not impossible to replicate.

Each utterance having unique F0s and their variations further supports the fact that in linguistics, F0 is perceived not in absolute, but in relative terms. However, as Zhang et al. (2017, 38) point out in their research, each phonological category, including phonetic segments such as vowels and consonants, and suprasegmental features such as lexical tones, possibly has only one abstract mental representation. A single abstract mental representation for each phonological category would, however, contradict the relativity of speakers' F0 in speech, as it would require an exact acoustic representation of each vowel, consonant, or lexical tone, including the intensity, duration, F0, and voice quality, each time it is uttered, otherwise it would not be successfully comprehended by the listener (Zhang et al. 2017, 38). Therefore, to ensure successful comprehension during day-to-day language use, it is theorised that a process called "perceptual normalisation" is subconsciously used by listeners at all times when exposed to spoken language. This process is thought to allow listeners to successfully classify a potentially infinite number of variables of acoustic realisations of different phonological categories (Zhang et al. 2017). In other words, even if listeners have a single abstract mental representation for each phonological category, they are, to a certain extent, capable of bending and adjusting the boundaries of each different acoustic representation and that way manage to retrieve the intended meaning successfully. However, such bending is, of course, not limitless. Very poor acoustic realisations may not be successfully classified by the listeners and therefore may lead to misunderstanding of the intended meaning.

1.1.2 Pitch and F0 in Music

As I have already hinted at the beginning of this chapter, both speech and music have some sort of structuring. As Trainor and Trehub (1992, 395) note, over time, in the same manner as language develops, changes, and refines its own conventional rules, structures, and systems, music evolves too. In Western music, there are conventional prototypical pitch relations among the set of tones which, in this case, is made of twelve semitones together forming an octave. The different combinations possible to be created from these tones then create expectations about which notes will follow (Trainor and Trehub 1992, 395–396). Moreover, some structures are then associated with evoking certain emotions. As already mentioned, the power to evoke strong emotions is shared by both music and speech. In music, this can be achieved through the semantic meaning of words, i.e., lyrics, if a piece of music contains them, but also through other aspects like rhythm, tempo, dynamics, timbre, and the pitch contours and pitch relations used in a song, all of which can be realised both vocally and instrumentally. In speech, this is primarily achieved vocally through prosody and the semantic meaning of the words spoken.

Importantly, as Zatorre and Baum point out (2012, 2), pitch in music is much more dependent on accuracy as, most of the time, it seeks the use of nearly exact values of F0 frequencies, as opposed to pitch in speech which, as already mentioned, is not interested in absolute values of the F0 frequencies speakers use, but it rather views those frequencies with respect to the speakers vocal range (Honorof and Whalen 2005, 2193). We can observe that in natural continuous speech, F0 changes constantly as there are no "steady-state pitches", whereas in music, a specific F0 is held for a particular amount of time, and then it changes and moves on to the next required F0 which is, again, held for a specific amount of time (Ladefoged and Johnson 2014, 126). Zatorre and Baum (2012) state that "…the concept of 'out of tune' does not even really apply to speech" (p. 2), whereas in music, even the smallest deviations from the expected frequency may be "…readily perceived as errors by listeners…" (p. 2). In contrast, in speech, such small deviations would not be perceived as errors at all. This difference exists possibly because in music, the phenomena of tonality and scales are used in most cultures, creating specific patterns, complex structures, relationships, and scales that are expected to be followed, however, no such phenomena are existent in speech (Zatorre and Baum 2012, 2; Trainor and Trehub 1992, 395).

1.2 The Development of Pitch Perception in Language and Music

As I will discuss momentarily, our native language, as well as music, affect us already prenatally (Mampe et al. 2009). It is therefore important not to overlook the development of these domains. This chapter targets exactly this issue with special (but not exclusive) focus on pitch perception, as this process is, for obvious reasons, prevalent throughout the first months of human lives.

1.2.1 The Development and the Early Role of Pitch Perception in Speech

Although this paper seeks to explore the potential language-to-music transfer based on the suprasegmental features of the two non-tonal languages (Czech and English), this subchapter discusses the segmental features of speech as well to provide a full picture of how pitch perception of speech develops.

As already discussed, in speech, pitch and its variations, together with prosody, are often used to convey lexical meaning, as is the case in tonal languages, but also grammatical meaning such as differentiating between a statement and a question, and extralinguistic meaning such as the speaker's attitude and affect (Liu et al. 2023, 1917; Zatorre and Baum 2012; Volín et al. 2015; Ohala 1984). Since prosody has such an important role in human communication, and since it is available to fetuses even before birth, its acquisition can begin as soon as the fetuses' audition and cognitive capacity has developed sufficiently, i.e. already prenatally. According to Mampe et al. (2009), prosody is an important feature of speech for infants acquiring their native language and because prosodic features of speech are, unlike segmental phonetic aspects, not disrupted by the abdominal barrier of the mother, they are even more distinct for human fetuses, making them easier to perceive. May et al. (2018, 8) share the same view, proposing that the most salient throughout the entirety of pregnancy are the prosodic properties of speech, as opposed to the significantly less salient segmental phonetic features. Chládková and Paillereau (2020, 1165) add that it is only the temporal information of speech that is transmitted to the fetus in an unchanged way. The transmission of spectral information, on the other hand, is notably limited. It is because of that, they add, that fetuses begin to learn durational characteristics and F0 contours earlier than features and contrasts based on higher frequencies that are not available auditorily inside the womb (Chládková and Paillereau 2020, 1165). Additionally, Zatorre and Baum (2012) point out that it has been shown that “contour information is more perceptually salient […] and more easily remembered, whereas specific intervals take more time to encode” (p. 3). On the same note, Mampe et al. (2009), discuss that “[h]uman fetuses are able to memorise auditory stimuli from the external world by the last trimester of pregnancy with a particular sensitivity to melody contour in both music and language” (p. 1994), in other words, they are already prenatally developing their “coarse-grained” perception which, as has already been mentioned, seems to be shared between both speech and music (Zatorre and Baum 2012; Liu et al. 2023).

Some studies (see Zatorre and Baum 2012; Mampe et al. 2009; Trainor and Trehub 1992) suggest that the consequences of the easier prenatal access to prosodic features than to segmental features, together with the perceptually more salient nature of F0 contours, are notably persistent throughout the first year of human life. As Zatorre and Baum (2012) argue, “[i]nfants detect contour but not interval information […], implying that it is a more basic process that develops early or is innate” (p. 3). Similarly, in their study, Mampe et al. (2009) examined the melodies of newborns’ crying. Based on their results which show that the melody contours of the newborns’ crying correspond to the intonation patterns of the newborns’ ambient languages, the authors concluded that fetuses start learning the prosodic aspects of their native language already prenatally. Further commenting on their experiment, they note that the reason for achieving such results may be that newborns show preference for their mother’s voice over the voices of others and that they are likely “…highly motivated to imitate their mother’s behaviour in order to attract her and hence to foster bonding…” and it seems that pitch contours “may be the only aspect of their mother’s speech that newborns are able to imitate…” (Mampe et al. 2009, 1994–1996; see also Weiss et al. 2015).

The conclusion that newborns prefer the voice of their mother is supported by Chládková and Paillereau (2020, 1159) who expand on it by noting that such preference is shown even by fetuses at least one month before the term. They add that newborns, as well as fetuses in the just mentioned stage of gestation, also exhibit preference for their native language. Finally, they (Chládková and Paillereau 2020, 1165) provide further insights into prenatal development of speech perception, proposing that it could either involve the process of “desensitization”, or the process of “attunement”. In other words, they describe that it is likely that fetuses either start becoming insensitive to foreign speech sounds that do not occur in their environment, or that they begin to attune to those sounds that are frequently present in their environment, leading to a significant preference for their native language sounds. They note that these processes may even operate jointly (Chládková and Paillereau 2020). As a result, newborns seem to "…display language-specific neural attunement, differentially processing their native language as opposed to nonnative-language (or otherwise unfamiliar) speech…" (Chládková and Paillereau 2020, 1159).

Trainor and Trehub (1992, 394) point out that speech pitch contours seem to be exceptionally salient for infants. Already at the age of 1 month, but perhaps even earlier, they are capable of discriminating between adult-directed and infant-directed speech with special preference for the latter. The two types of speech, they add, differ noticeably from one another in the rich presence of particularly those features which are typically associated with music rather than speech, such as greater melodicity, wider F0 contours, and slower articulation. What stands out the most to infants, underlying their preference for infant-directed speech, is the pitch contours (Trainor and Trehub 1992, 394). Trainor and Trehub (1992, 394) also note that infants exhibit more positive affective displays when they are exposed to such pitch contours that are associated with maternal utterances of approval rather than those that are associated with utterances of prohibition (Trainor and Trehub 1992, 394; see also Weiss et al. 2015). They (Trainor and Trehub 1992) also state that "…infants 5 to 11 months of age readily discriminate tone sequences, or tunes, differing in pitch contour" (p. 394), and that they are even capable of categorizing and grouping these sequences on the basis of their either identical or even just similar pitch contours, supporting the prevalence of prosody over phonetic segments or individual tone intervals.

Based on this prevalence of prosody, especially that of pitch contours, Trainor and Trehub (1992) offer a hierarchy of features of infant speech perception, suggesting that infants are initially sensitive to suprasegmental or musical aspects of speech, especially pitch contour, and although they do exhibit some sensitivity to segmental aspects of speech as well, effects of experience on phonetic perception are not apparent until 10–12 months of age (Trainor and Trehub 1992, 394). In summary, they suggest that for processing of speech sounds, infants rely more on the acoustic and phonetic features of speech rather than the phonetic categories and the speech sounds themselves, although, that does not mean they do not develop these categories. According to Chládková and Paillereau's review (2020, 1137), infants create the perceptual categories for vowels of their native language sometime between the 4th and 6th month of age and the categories for consonants of their L1 between the 10th and 12th month.

Additionally, Trainor and Trehub (1992) note that infants are much better at differentiating foreign speech sounds from languages that are nonnative for them than their parents. This is because infants still utilize pre-phonological processing, as Trainor and Trehub call it, which allows them to differentiate members of native contrasts and of nonnative contrasts with equal ease. Adults, on the other hand, together with older infants as well, as I will mention momentarily, utilize phonological processing of speech which is what makes them favour native language speech sounds over speech sounds from nonnative languages, resulting in reduced or no perceptual sensitivity to the nonnative sounds (Trainor and Trehub 1992; Chládková and Paillereau 2020, 1137). The change from the pre-phonological processing to the phonological one, as they conclude, takes place around the infants' first year of life. At that point, this perceptual hierarchy changes, assimilates to, and eventually matches that of an adult, resulting in the loss of the discussed advantage of infants in differentiating foreign speech sound contrasts (Trainor and Trehub 1992). This phenomenon is known among developmental speech researchers as "universal listener".

"Universal listener" is a term used for infants that do not yet employ the phonological processing but are still using the pre-phonological one which seems to be present since birth and lasts only, if at all, a few months (Chládková and Paillereau 2020). The ability to distinguish any nonnative speech sounds, even those the infants have never heard before, is why they are seen, as Chládková and Paillereau (2020) put it, as "…*universal listener*[*s*] with *universal perceptual abilities*" (p. 1137). However, it is important to note that opinions differ on the phenomenon of "universal listener". Linguists are not in agreement as to when exactly and for how long humans exhibit these universal perceptual abilities, or whether such perception even exists at all. Based on previous research, Chládková and Paillereau (2020, 1159) suggest that when it comes to suprasegmental aspects of speech, newborns show signs of language-specific speech perception. However, little is known about newborns' language-specific perception of individual segmental speech sound properties. They come to the conclusion that humans probably no longer display entirely language-general, i.e., universal, speech perception abilities at birth but only somewhere between the 28th week of gestation and birth (Chládková and Paillereau 2020, 1170). Linguists that support the theory of the universal listener argue that this ability is what gives infants the predispositions to acquire virtually any natural language as their native one.

Finally, although children are capable of producing speech sounds in an adult-like fashion around the age of 5, they acquire the specific perceptual and comprehension abilities needed for their native language much earlier. In fact, infants seem to develop substantial knowledge about the speech sounds that make up their native language already before the 7th month of life when they start to babble (Chládková and Paillereau 2020, 1137). However, despite the temporal difference in when speech perception and production by infants become observable, the processes are probably inherently interconnected and therefore need to be treated accordingly, i.e. it is crucial to explore one with the context of the other (Casserly and Pisoni 2010). This is supported by Akahane-Yamada et al.'s (1996) experiment. They tested young adult native Japanese speakers to identify English /r/-/l/ minimal pairs, i.e., segmental differences. Half of the participants received special perception training. The results show that speech perception training provides a long-term improvement in not only speech perception, but also in speech production.

1.2.2 The Development and the Early Role of Pitch Perception in Music

As already discussed, humans focus on musical as well as speech pitch contours prenatally (Mampe et al. 2009), beginning to develop their "coarse-grained" perception skills useful for melody discrimination, very early on in their life. Mampe et al. (2009, 1994) add that newborns even show preferences for melodies to which they were exposed prenatally. As Weiss et al. (2015) suggest, there are various differences across age groups of young children (pre-teenage) in their abilities to recognize old and new melodies. Their results show that for 7-year-olds and older, vocal melodies seem to be more easily discernible than instrumental ones. Additionally, they note that melody memory improves noticeably with age. Although they stress that further research with greater variability of stimuli is needed, their experiment shows that children only two years older than their younger co-participants, namely 11-year-olds as compared to 9-year-olds, exhibit better memory for melodies, allowing them to differentiate between old and new melodies with even better precision. On the same note, Weiss et al. (2016) tested the degree of pupil dilation of fifty young adults, roughly of the age of 20, when exposed to different combinations of vocal or instrumental and familiar or unfamiliar melodies. Based on the results, they suggest that the exhibited greater pupil dilation for familiar vocal melodies is a result of listeners' heightened arousal for or engagement with those melodies, adding that these results provide further evidence for the special importance that vocal music has for humans (Weiss et al. 2016, 1063).

Music plays a particularly important role in the upbringing of a child. According to Cirelli and Trehub (2020) who examined 8- and 10-month-old infants, singing familiar songs to infants in distress soothes them more effectively than singing unfamiliar songs and even more effectively than trying to sooth them by expressive speech. Interestingly, according to Trehub et al.'s research (1997), adults, both male and female, with and without experience in childcare, and also with and without musical training, seem to produce different results when asked to simulate infant-directed singing as opposed to when asked to actually sing to an infant. The real infant-directed singing differed in many aspects as it involved emotional expressiveness and engagement, slower tempo, and higher pitch levels, which are often connected with happiness, tenderness and affection (Trehub et al. 1997). Trehub et al. (1997) comment on these results by noting that the emotional expressiveness is most likely heightened due to the biological need for bonding which can be achieved through emotional engagement. Adults incline to slower tempo probably because they want to adjust to the limited processing capacity of infants. The combination of slow tempo and higher pitch levels is naturally used by adults as it seems to enhance the soothing effect a song may have on an infant (Trehub et al. 1997, 505).

As already discussed, apart from "coarse-grained" perception, music also utilizes "fine-grained" perception which seems to be unique to music (Zatorre and Baum 2012; Bidelman et al. 2011a; Liu et al. 2023). As Zatorre and Baum (2012) suggest, the "fine-grained" perception focusing on "precise encoding and production required for musical scale information might be a separate mechanism, perhaps even one that emerge[s] later in phylogeny" (p. 3). Last but not least, it is important to mention that poor pitch production during singing is not necessarily due to deficits only in the production abilities or only in the perception abilities of the individual. The deficits seem to lie somewhere between the two (Pfordresher and Mantell 2009, 425).

1.3 Cognitive Processing of Speech and Music

Since this thesis is concerned with a potential language-to-music transfer, it is important to review whether or not speech and music are somehow related on the level of cognition. As already briefly hinted when discussing the difference between the coarse- and the fine-grained perception (for more information, see Zatorre and Baum 2012 and Liu et al. 2023), the cognitive processes underlying speech and music are to a certain extent intertwined, allowing for potential language-to-music transfer. Many studies argue for and/or provide evidence that a link between the brain regions connected with speech and music processing exists (see Casserly and Pisoni 2010; Hutka et al. 2015; Mantell and Pfordresher 2013; Besson et al. 2011; Maess et al. 2001; Koelsch et al. 2002; Slevc et al. 2009; Bidelman et al. 2011b; Bidelman et al. 2013; Deutsch et al. 2011; Asaridou and McQueen 2013; Patel 2011; Warren 2008; Akahane-Yamada et al. 1996). Hutka et al. (2015) provide strong evidence by reviewing previous literature concerned with the topic of cognitive processing of speech and music. They say that "…the processing of musical melody and harmony activate brain areas traditionally associated with language-specific processing, such as Broca's and Wernicke's areas…" (Hutka et al. 2015, p. 52). Similarly, Marques et al. (2007) point out that Broca's area, a key region for language processing, "…has been shown to be activated not only by syntactic processing of linguistic phrases but also by syntactic processing of musical phrases" (p. 1453). Mantell and Pfordresher (2013, 199) provide further evidence for the interconnectedness of the cognitive processing responsible for speech and music. They found that imitation abilities in one domain (e.g., speech) predict these abilities in the other domain (e.g., music) as well.

Interestingly, musicianship plays a more important role in the cognitive processing of speech and music than may be expected. Musicianship, not only professional, but amateur as well, has been shown to influence both morphological organisation and functional activation of the brain. In other words, musicians have been observed to have morphological differences in some brain areas as opposed to non-musicians, and also to show larger activation in several brain regions (Marques et al. 2007, 1453). Importantly, though, Marques et al. (2007, 1453–4) add that the regions in question do not serve exclusively for the processing of music, but also for other types of perceptual and cognitive processing, including language processing. They conclude by hypothesising that musical experience may therefore facilitate language processing as it increases activation even in those brain regions that are, at least to a certain extent, responsible for this processing.

It now becomes more evident that the related neural mechanisms responsible for the cognitive processing of speech and music are nearly inseparable. The close connection and overlaps between the related brain regions make setting strict boundaries between the cognitive processing of speech and the cognitive processing of music near impossible. However, such is not the aim of this paper anyway. The discussed findings provide significant evidence for the interconnectedness of the cognitive processes underlying speech and music processing, allowing for the here-examined language-to-music transfer.

1.4 F0 Contours

As Volín et al. (2015, 109) point out, despite the absence of any physiological differences, different linguistic communities tend to use various pitch ranges in their speech. Skarnitzl and Hledíková (2022, 1) add that the intonation of the Czech language is significantly flatter than that of English. This could partly be because prosodic phrases of Czech are longer than those of English by nearly 40 % which results in fewer long stretches of speech which lack salient melodic movements (Skarnitzl and Hledíková 2022, 4). Based on the data, they briefly describe Czech prosodic patterns as monotonous and the English ones as vivid. Therefore, now that I have established the importance and the development of pitch and F0 in speech and music, it is only appropriate to investigate in greater detail the average width of the F0 contours of English and Czech, i.e., the average F0 range that is used by native speakers of these languages in day-to-day speech, as well as the intonation patterns of these languages.

First and foremost, it is important to note that while both these languages use F0 contours for similar purposes (as already discussed in chapter 1.1.1), English seems to exploit intonation contours slightly more than Czech as it also relies on them to express contrast or to stress important information. Czech, on the other hand, achieves these purposes through the means of grammatical inflection or word order (Volín et al. 2015, 107). This slight but not negligible difference gives English native speakers yet another opportunity to further refine their perception and production skills which may result in their potentially greater abilities in musical pitch perception and production. As Volín et al. (2015) point out, an alarming portion of foreign speakers of English, including the Czechs, are completely unaware of this additional use of intonation in English. This is, possibly, what contributes to the widespread foreign-accented Czech English which is common even among more experienced Czech users of English. However, overlooking this extra employment of intonation in English does not merely lead to the production of foreign-accented English, but it may lead to misunderstanding of the intended meaning or even to complete breakdown of the communication (Volín et al. 2015, 107–108).

Non-native production of language has many typical features, one of which is narrower F0 span (Volín et al. 2015, 109). This could be the result of potential uncertainty or anxiety that the speakers may feel when they must speak a foreign language (Volín et al. 2015, 121; see also Volín et al. 2017). Additionally, it has been observed that foreign-accented speech is heavily influenced by the speaker's mother tongue (Volín et al. 2015, 121). In the case of Czech English, this typically results in what sounds as flat, monotonous language, resembling boredom, or lack of interest or involvement (Volín et al. 2015, 109). On the same note, Volín et al. (2017) examined the speech of sixteen American and English speakers of Czech who were asked to read several paragraphs from Czech news broadcasts. In accordance with Volín et al.'s 2015 study, the participants exhibited narrower pitch range than that of native English, and often even narrower than that of native Czech. This provides further evidence that non-native speakers of a language often use much narrower F0 range, possibly due to the mentioned uncertainty of speaking a different language. Narrower pitch span, in turn, results in what is perceived as foreign-accented language production. To eliminate this problem, specific focus on phonetics and intonation of English in (not exclusively) Czech education systems is strongly advised for by Volín et al. (2015, 108), who observe previous studies and point out that native-like prosody is what, according to native speakers, highly enhances the intelligibility of non-native speech production of their native language.

More importantly, Volín et al. (2015, 113–114) studied the average level of F0 employed in Czech and English and found a significant difference between 32 native English and Czech professional newsreaders. According to their results, the Czech speakers used averagely lower F0 frequencies in their speech as both the mean and the median were observed to be of higher frequencies in the collected English dataset as opposed to the Czech one. This trend, however, disappeared when randomly chosen non-professional speakers were examined in addition to compare the results. As they point out, this could be an indicator that BBC either requires higher involvement from their newsreaders, as higher involvement entails the use of higher pitch, or that they hire only people with higher-pitched voices. Most importantly, though, Volín et al. (2015) studied the average F0 range used by these speakers, closely accounting for all deviations which could manipulate the results. They examined not only the potentially problematic variation range which relies solely on the highest and the lowest extremes, but also the range between the 10th and the 90th percentile, the range between the 1st and the 3rd quartile, and finally also the standard deviation. All four of these metrics provided evidence that the English speakers deviated from their average pitch level significantly more than their Czech counterparts and that they did so in both the directions, i.e., they produced both higher and lower frequencies than the Czech speakers. In summary, they discovered that English speakers utilise a wider F0 span than Czech speakers. The difference between the two languages, as their results show, does vary depending on different speech contexts with different communicative functions. However, although Czech speakers may come closer to the average level of F0 contours of English in some situations, on average, the English speakers still utilise wider F0 span in their speech (Volín et al. 2015; see also Skarnitzl and Hledíková 2022).

These findings were further extended by Skarnitzl and Hledíková (2022), who examined twenty skilled native Czech and (American) English TED Talk speakers, ten of each language, who were chosen for their high-quality delivery. Their findings align with those of Volín et al. (2015), providing more evidence that melodic variations and average F0 deviations of Czech speakers are noticeably narrower than those of English speakers. Interestingly, the results of Veronika Vonzová (2023) show that British native speakers use even wider F0 range than American speakers in the majority of the intonation patterns, the only exception being the rise-fall movement. The results of Skarnitzl and Hledíková (2022) that suggest that Czech native speakers employ narrower pitch span than American native speakers of English have therefore even higher significance as this means that the difference in the F0 span employed by Czech native speakers and British native speakers of English should be even greater. Last but not least, further data from Skarnitzl and Hledíková (2022) show that while prosodic phrase length seems to be, apart from language-specific, also genre-specific, pitch span seems to be solely language-specific. More precisely, their participants, i.e. (American) English and Czech native TED Talk speakers, exhibited a similar width of pitch contours as the newsreaders examined by Volín et al. (2015), however, the length of the prosodic phrases was noticeably shorter among the TED Talk speakers, suggesting that genre plays an important role in this respect, and that it needs to be closely accounted for in further research in this field (Skarnitzl and Hledíková 2022).

As I have now mentioned several important terms related to the topic of F0 contours, such as prosodic phrases, intonation patterns, and rise-fall movement, it is only appropriate to describe these in further detail as these are integral to the herein-discussed concept of F0 contours, i.e., pitch contours, of a language. F0 contours, as intended by this thesis, and as has already been mentioned multiple times, denote the F0 span employed by speakers for their native language production. In this general concept, F0 contours do not have any horizontal boundaries, i.e., they are not segmented or structured. This thesis views them as a whole, as a construct that does not need to be divided at all. Contrarily, prosodic phrases, intonational phrases and patterns, and intonation movements do have set boundaries based on which they can be examined.

As Wells (2014, 103) defines, intonation is a complex process which consists of, as he calls them, the "three Ts", i.e., the systems of *tonality*, *tone*, and *tonicity*. Wells describes these as the three types of decisions speakers have to constantly make when speaking a language. Respectively, these decisions are: (1) chunking the speech into segments (i.e., signalling the syntactic boundaries), (2) choosing which pitch patterns to use, and finally (3) choosing where to put accents in the utterance. Ladefoged and Johnson (2014, 25) provide a briefer definition, saying that intonation is the pitch pattern of a sentence. As they add, intonation segmented into smaller chunks than a sentence is then regarded as an intonational phrase. Each intonational phrase consists of an intonation pattern which usually extends over the entirety of the intonational phrase. As Wells (2006, 187) describes, it is not always easy to distinguish between individual intonational phrases. Most of the time it is a matter of common sense and feel for the language, however, intonational phrases usually mirror grammatical structures, meaning that an intonation break, which is the boundary between two successive intonational phrases, usually occurs together with a syntactic, i.e., grammatical, boundary. Intonation breaks therefore tend to occur between successive sentences, clauses, phrases, and occasionally also words. Lastly, prosodic phrases are similar to intonational phrases, however, as described by Frazier et al. (2006, 244), apart from intonation patterns, prosodic phrases also include the other aspects of prosody such as rhythm and stress properties.

Finally, it is appropriate to describe the intonation patterns and movements, of English and Czech as, essentially, it is them that create the difference in the overall average F0 range employed by native speakers of these languages, and it is exactly this difference that prompted my idea to delve into this topic. Intonation patterns which, as already described, form intonational phrases, consist of intonational movements which are the specific pitch movements we can observe within an intonational phrase. There are two main traditions for studying intonation of English – the British tradition, and the American tradition (Vonzová 2023, 34). For my purposes, I will refer to and use the terminology of the British tradition, which views intonation in terms of contours that involve pitch movements and nuclear stress (Vonzová 2023, 37). Levis and Wichmann (2015, 140) describe that British tradition segments intonation into "tone groups" which may form a complex tone group structure (for further information, see Levis and Wichmann 2015), however, the only essential element that cannot be omitted from the structure is the *nucleus*. They describe *nucleus* as an accented syllable which carries pitch movement. Such movement is known as *nuclear tone*. In case there is more than one syllable of such kind within a single intonational phrase, it is the last one which is the *nucleus*. The *nuclear tones*, i.e. intonation movements, can be described as follows: *fall, rise, fall-rise, rise-fall,* and *level*.

Ladefoged and Johnson (2014, 134) point out that intonation cannot be easily described as it varies among different varieties of English, among individual speakers, and what is more, it differs even based on individual occasions of utterance and individual nuances of meaning. Despite the difficulty of generalising the rules for the use of intonation, they provide a brief general overview of the most common intonation movements and their usage in standard varieties of English. In declarative sentences, there is a tendency to end with a falling pitch, i.e., *fall movement*, unless there is another intonational phrase to follow within the same sentence. In such a case, a movement described as *continuation rise*, or at least a lack of the final *fall*, can be observed. Questions asking for yes or no answers usually end with the *rise* or the *rising movement*. The *rise movement* is larger, more prominent, than the *continuation rise*. Finally, they conclude that the so-called "wh- questions" usually end with a *fall*. Additionally, Levis and Wichmann (2015, 149) broadly describe that rising contours are usually used to indicate openness or non-finality. Falling tones, on the other hand, indicate closure or finality. Their further observation of the usual usage of the specific intonation movements is in absolute agreement with Ladefoged and Johnson's (2014) description I just provided as they note that "…statements generally end low, questions often end high, and […] nonfinal tone groups in a longer utterance also frequently end high, signaling that there is more to come" (Levis and

Wichmann 2015, 149). They add that question tags assuming confirmation appear with a *falling* tone, and question tags seeking information with a *rising* tone. This description is enough for my purposes of giving a brief overview of the intonation movements in English. For further data, see Vonzová (2023) who examined the intonation patterns of British and American English in great detail and compared the results (see also Ladefoged and Johnson 2014; Levis and Wichmann 2015; Wells 2006, 2014; Chamonikolasová 2017).

Describing Czech intonation is much more difficult as there is significantly less data and studies concerned with this topic available. Laštůvka (2023) reviews relevant literature, observing that the structure of Czech intonation is very similar to that of English as it also requires the presence of pauses for separating the individual intonational phrases, and it also tends to reflect syntactic structures. These two properties jointly indicate the boundaries between the individual segments. Finally, the individual structures are also built around prominent peaks which can be equated to English *nuclear tones* (Laštůvka 2023, 16). The Czech near equivalent of the English intonational phrase is referred to in Palková's Czech terminology as *promlouvý úsek* (Palková 1994, 162). Although she notes that there is no formally established English equivalent for the term (at least at the time of her writing her book), I believe that for the purposes of my thesis, these can be viewed as equal as their structuring and behaviour are virtually identical. She further describes *promlouvý úsek* as a single intonational unit with relatively clear boundaries which is perceived by language users as an inherent whole. Chamonikolasová (2017, 18) adds that each *utterance unit* (an English translation used by her to refer to the Czech term *promlouvý úsek*) contains a *melodeem* (*melodém* in Czech) which is the most prominent pitch movement of an *utterance unit*, usually occurring at its very end. Once again, I believe that for my purposes, Czech *melodeem* can be broadly equated to English *nuclear tone* as it also describes its associated pitch movement (Laštůvka 2023, 17).

Finally, Laštůvka (2023, 17–18) provides a useful brief overview of the three main intonation movements of Czech, adding his own translations of each for more clarity. *Melodém ukončující klesavý* (translated by Laštůvka as *final falling contour*) is the movement that occurs most frequently in Czech, and it correlates to the English *fall movement*. Its distribution is also the same as that of its English counterpart, as it is used at the end of declarative sentences. *Melodém ukončující stoupavý* (translated as *final rising contour*) serves as the main indicator of questions since Czech, unlike English, does not rely on word order in this respect. Finally, *melodém neukončující* (translated as *non-final contour*) is the only Czech intonation movement which does not occur at the end of an utterance for it signals continuation (similarly to English *continuation rise* described above). This movement therefore typically occurs in complex and compound sentences, and its pitch pattern itself can be either slightly *rising* or slightly *falling* (Laštůvka 2023, 18). For further information on Czech intonation, see the discussed works (Laštůvka 2023; Palková 1994; Chamonikolasová 2017).

1.5 Factors Relevant to the Experiment

The entirety of this chapter is dedicated to the variables related to both the participants and the stimuli of the experiments I will propose in chapter 2: “Methodology”. It will focus in greater detail on the individual aspects that need to be accounted for when testing musical pitch perception and production.

1.5.1 The Effect of Musicianship on Musical Pitch Skills

It is common sense to expect that musicianship, musical education, and intense musical experience have a significant effect on one’s musical abilities. In this section, I will focus on this topic in greater detail.

Marques et al. (2007) separated French adults into musicians and non-musicians and presented them with sentences spoken in Portuguese. The final words of the sentences were either unaltered, or their pitch was increased by either 35 % or 125 %. Their results show that musicianship enhances the processing of fine pitch variations, i.e., musicians better discriminated the stimuli with pitch altered by 35 %. Finally, the authors point out that since musical expertise enhances the ability to discriminate pitch, it should facilitate pitch variation processing in not only speech but in music as well. Similarly, Wayland et al. (2010) suggest that musicians have more refined pitch perception abilities. They add that intensive pitch perception training improves pitch perception abilities notably not only among non-musicians, but among musicians as well. Furthermore, extensive musical training seems to enhance the ability to distinguish “…nuances in emotional expressiveness” (Trehub et al. 1997, 506). Finally, Montagni and Peru (2012, 33) add that pitch discrimination is strongly enhanced not only by regular music practice, but also by early exposure to foreign languages. Deutsch et al. (2004b, 370) provide very similar results, specifying that exposure to different foreign speech sounds in early life can heavily influence one’s perception of musical patterns.

As already discussed, there is evidence that tone L1 background influences one’s musical abilities (see Liu et al. 2023; Choi 2021; Chen et al. 2016; Ngo et al. 2016; Swaminathan et al. 2021; Bidelman et al. 2013; Zhang et al. 2020; Wong et al. 2012). Interestingly, though, based on their research, Hutka et al. (2015) and Choi (2021), who provide identical results, suggest that musicianship enhances one’s musical pitch perception far more than tone L1 background. Similarly, Cooper and Wang (2012) provide evidence that musical experience facilitates tone word identification to a larger extent than tone-language background does. In agreement with Cooper and Wang (2012), the data of Alexander et al. (2005) show that musicians are better than non-musicians at discriminating the lexical tones of a tone language, Mandarin Chinese.

Choi (2021) further expands on his results, noting that L1 background provides an advantage for musical pitch perception only among non-musicians, suggesting that musicianship overrides any potential L1 background advantage. His results, although related partly to linguistic rather than strictly musical abilities, are still relevant since the domains for music perception and production are, as already discussed, probably inseparably associated, potentially forming a bi-directional relationship (Cooper and Wang 2012; Hutka et al. 2015; Choi 2021; Alexander et al. 2005; Wayland et al. 2010; Montagni and Peru 2012).

1.5.2 The Effect of Absolute Pitch on Musical Pitch Skills

Absolute pitch, also known as perfect pitch, is "…the ability to identify the pitch of a musical note or to produce a musical note at a given pitch without the use of an external reference pitch" (Takeuchi and Hulse 1993, 345; see also Sergeant 1969; Deutsch 2002; Deutsch et al. 2004a; Deutsch 2013). Most people, when processing musical pitch, focus on the relationships among the individual notes rather than observing each note in isolation. In other words, they perceive musical pitch relatively, i.e., with respect to the neighbouring tones, rather than absolutely, i.e., perceiving the absolute F0 value of each tone. It is exactly in this respect that people with absolute pitch differ from the rest. They are able to encode and remember the absolute pitch values (Takeuchi and Hulse 1993, 345), although, as Levitin and Rogers (2005, 28) stress, the possessors of absolute pitch are not always 100 % precise and make errors too.

Deutsch et al. (2006) note that the incidence of absolute pitch is extremely rare in the U.S. and Europe, estimating that, on average, less than one person in 10,000 can pride with this ability. Despite that, though, Deutsch (2002) notes that, at birth, the ability to acquire absolute pitch may actually be universal for speakers of all languages, and that it can be "…realized by giving the infant opportunity to associate pitches with verbal labels during the 1st year or so of life" (p. 200). Her findings are supported by Deutsch et al. (2006) who also add that the sooner one starts musical training, the greater is the chance for them to acquire absolute pitch. Deutsch et al. (2009) specify that absolute pitch is most prevalent among people who have already started musical training by the age of four or five, and that it is quite rare among those who have started musical education after the age of nine. Finally, they revise previous findings, noting that any attempts to train adult musicians to acquire absolute pitch were unsuccessful.

Interestingly, overall higher prevalence of absolute pitch was found among Asian music conservatory students (Deutsch et al. 2006). Deutsch et al. (2009) provide similar results. They found that students of East Asian ethnicity who spoke tone language fluently performed better on a test for absolute pitch than those who were less fluent at speaking a tone language and even more so than those who did not speak a tone language at all. It is important to point out that despite the prevalence of absolute pitch among the Asian population, and even higher prevalence among those speaking a tone language, absolute pitch is by no means a pre-requisite to acquire, speak, or understand any natural language. Finally, it is important not to confuse absolute pitch with the ability known as "relative pitch". Although they may seem similar, they are significantly distinct skills. Relative pitch is the ability to identify and produce the intervals, i.e., the relationships, between individual notes. This ability, unlike absolute pitch, can even be learned through regular music practice, and as a result, most professional musicians possess it (Levitin and Rogers 2005). Last but not least, there is some evidence that a unique type of absolute pitch, specifically on the level of production, is exhibited by native speakers of tone languages in speech as they seem to enunciate the same tone words with the same absolute F0 patterns even when being tested on different days (Deutsch et al. 2004a).

1.5.3 The Properties of the Stimuli of the Experiment

To choose adequate stimuli for the experiments, it is important to know the different options there are when it comes to different types of sounds. Crystal (2008) defines *tone* in acoustic phonetics as "…a sound with sufficient regularity of vibration to provide a sensation of pitch" (p. 486). A sound without the periodic vibrations is known as *noise*. *Noise*, however, is not a useful sound for testing musical pitch perception and production as we do not perceive its pitch due to its lack of regular vibrations. A *pure tone* is a sound "…whose pattern of vibration repeats itself at a constant rate" (Crystal 2008, 486). Such sounds, however, do not often occur naturally but are usually produced by electronic sources. Finally, *complex tone*, or acoustically rich tone, is a result of a combination of two or more tones of different frequencies. These are the sounds we come across most often as they occur, e.g., in natural speech (Crystal 2008, 486).

Iino et al. (2018) tested whether humans elicit different mismatch negativity (MMN) response to pure tones as opposed to spectrally rich speech sounds, i.e., complex tones. Their findings show that MMN response was larger as well as faster to speech sounds as opposed to pure tones. They propose that this could be because humans are accustomed to hearing speech sounds on daily basis whereas pure tones are much rarer, therefore possibly harder to decode. They support their reasoning with the findings of Pulvermüller and Shtyrov (2006) who found that subjects elicit larger MMN response to familiar sounds over unfamiliar sounds. Additionally, Iino et al. (2018) reveal that with age, the amplitude of MMN decreases and the latency increases for pure tone sounds. Importantly, they observed no such effects of age for speech sounds. These findings provide significant insights into the creation of adequate stimuli for the experiments. Acoustically rich speech sounds appear to be a more reliable type of stimuli as they are more easily and readily perceived by all speakers, and they do not carry the need to control for the participants' age.

Weiss et al. (2012) examined whether or not adults remember vocal melodies better than instrumental ones. Their results provide evidence that adults indeed remember vocal melodies with better precision than they remember instrumental melodies. Their related study (Weiss et al. 2015) shows that children exhibit perceptual preference for vocal melodies over instrumental melodies, although the preference is represented differently across the different age groups. This data suggests that voice is more salient to humans than instruments. Weiss et al. (2012) suggest that this may be because speech is biologically significant to humans. This may project into vocal music, leading to "…increased vigilance or arousal, which in turn may result in greater depth of processing and enhanced memory for musical details" (p. 1074). It would therefore be ideal, to account for this by not randomly mixing vocal and instrumental sound stimuli, but rather putting the two different types of stimuli together systematically. Alternatively, separate experiments, each using a different type of stimuli, can also help avoid potentially biased data.

According to Trainor and Trehub (1992), different cultures have different traditions of music and therefore have different expectations of what a "good melody" is. Generally speaking, though, a melody is considered good or well-structured by a person as long as it conforms to the rules and norms of music of their culture. Trainor and Trehub (1992) note that "[t]here is some evidence that 8- to 10-month-olds show enhanced processing for certain musically well-structured, or good, melodies" (p. 395), adding that "…adults without formal musical training are sensitive to this structure [as well], being better at processing melodies that exemplify the musical structure of their culture than melodies that do not…" (p. 395). As already hinted in chapter 1.1.2, Western music is characteristic for its use of diatonic musical scales which consist of unequally distanced tones. Eastern music, on the other hand, typically uses chromatic scales which are characteristic for their use of equally distanced tones. Trehub et al. (1999) support Trainor and Trehub's insights (1992) by providing evidence that western infants are better at detecting subtle pitch changes in diatonic scales, rather than in chromatic scales. Similarly, adults detected subtle pitch changes in familiar diatonic scales with better precision than in unfamiliar diatonic and chromatic scales. When testing for musical pitch perception and production abilities, the ethnic background of the participants therefore needs to be considered when choosing the stimuli.

Finally, Peter et al. (2014) asked children and adults to vocally imitate nonwords spoken by a man and a woman. They revealed that when the F0 of the target nonword was within the vocal range of the participants, they approximated that F0 level. Additionally, they provide the following results for situations when the target F0 was outside the participants' vocal range. When men were asked to replicate nonwords spoken by a woman, they approximated the F0 level one octave below the target F0. Children who were asked to replicate a man's voice approximated the F0 level one octave higher than the target value. Surprisingly, women imitating a man's voice approximated the target F0 "…at a ratio of 1.5 known as the perfect fifth in music" (Peter et al. 2014, 1). All the participants adjusted the target pitch as described according to their vocal range without being given specific instructions to do so. Each participant's approximate vocal range was measured prior to the experiment through the second edition of the Goldman-Fristoe Test of Articulation. This test presents participants with picture stimuli in order to elicit word productions that can be recorded and examined (Peter et al. 2014). These results suggest that in pitch production experiments, the target F0 value should be taken into account with respect to the participants' vocal range.

2 Methodology

Based on the literature reviewed in the previous chapter, I will now propose an adequate methodology for experiments testing musical pitch perception and production. I will also provide and comment on two already existing online experiments that test for musical pitch perception. Since it is not the aim of this study to carry out the experiments, collect the data, and comment on the results, it is encouraged to use or at least consult the methodology proposed here for any potential future experiments. Note that "fine-grained" perception is targeted in the experiments as well even though it does not seem to be influenced by native language background. It is included to potentially provide further evidence that "fine-grained" perception can truly only be enhanced through musical practice.

2.1 Testing Musical Pitch Perception

In this section, I will provide and comment on two already existing online experiments that can be used to test musical pitch perception. Then, I will propose adequate methodology for creating and carrying out a unique experiment targeting this issue.

2.1.1 The Existing Online Experiments

In case of any potential future interest in this topic, I will provide two existing online experiments that may be readily used to test musical pitch perception. These are: 1) Profile of Music Perception Skills (PROMS) available on musemap.org (Musemap 2024), and 2) a test of musical IQ (hereafter referred to as MIQ) available on themusiclab.org (The Music Lab 2024). Both the experiments test for "fine-" and "coarse-grained" perception, and each of them does so slightly differently.

MIQ (The Music Lab 2024), when testing for "coarse-grained" perception, presents melodies played on an electronic piano (or a synthesiser). The melody is presented three times, each time in a different musical key. One of the three versions has a mistake in the melody (i.e., some of the intervals are changed) and the participants are asked to choose the odd one out. When testing for "fine-grained" perception, subjects are presented with a section of a song that includes both instrumental parts and sung lyrics. Participants hear the melody twice and then are asked in which of the two was the singer more out of tune. In both parts of the experiment, some of the differences are very minor, and some are more evident. All the melodies conform to the traditions of Western music. MIQ also tests for perception of musical beat, however, that is not relevant to this topic.

PROMS (Musemap 2024) has multiple variants of the experiment which primarily differ in the size of the stimuli presented. The herein-discussed variant is the "Mini-PROMS". To test for "coarse-grained" perception, a melody played note by note on an electronic piano (or a synthesiser) with no other instruments or chords in the background is presented to the participants. This serves as "reference" stimuli. It is presented two times. The third time, a comparison sound is played. The comparison sound is either the same as the reference, or some intervals are changed. The participants are asked to choose whether the comparison sound was the same or different from the reference stimuli.

Testing for "fine-grained" perception, PROMS uses the same structure as when testing "coarse-grained" perception, presenting the participants with reference stimuli two times and then presenting them with a comparison sound. The stimuli are either chords played on an electronic piano (or a synthesiser), or they are single pure tones played in isolation. The comparison sound is either the same as the reference sound, or it differs from the reference sound by having one tone from the chord slightly out of tune. In the case of the pure tones, the comparison sound may differ by having its pitch slightly shifted. The participants are asked whether the comparison sound was the same or different from the reference sound. PROMS also tests for beat and tempo perception, however, as mentioned above, none of these are relevant to this study.

The following subchapters will focus on creating a separate unique experiment. The methods used will be inspired by the two discussed online experiments as they both test for musical pitch perception abilities differently, each using a different degree of complexity and variety of the stimuli.

2.1.2 Procedure

Testing musical pitch perception may be possible through both online and in-person experiments. Of course, in the case of the online experiment, it is impossible to control over the quality of the participants' headphones/speakers, and to ensure quiet environment during the testing. On the other hand, the online experiment allows for a much larger sample size, which may, in turn, substitute for the possible related deficiencies. It seems most practical to create a singular test in a computer format for both the online and the offline versions of the experiment. An introductory section stressing the importance of the usage of quality headphones as well as a quiet environment during the proceedings of the test may be added to the online version of the experiment to minimise the mentioned side effects. As for the offline experiment, the participants would be asked to come to a test room at a given location. The room should be silent and equipped with quality headphones and a computer on which the participants would complete the test. Otherwise, both the online and the offline experiments should proceed similarly.

Firstly, each participant should fill out a short questionnaire asking for their name, age, sex, nationality, L1, foreign language experience, musical experience, and whether they have any hearing or speaking disorders. Next, each participant's approximate vocal range may be tested, e.g., via the Goldman-Fristoe Test of Articulation, as done by Peter et al. (2014). The results of the experiment should then be compared with relation to the participant's average vocal range, considering whether the stimuli were or were not within their vocal range, to explain potential deviations and provide data for further related research. The rest of the procedure of this experiment proposal is inspired by the procedure of PROMS.

The test should consist of at least five parts, each differing in the type of the stimuli used. Participants will be asked to listen to the reference stimuli two times. Then, the comparison stimuli will be played, being either identical or different from the reference stimuli in a specific way as discussed in greater detail in the following subchapter. Finally, after each set of stimuli containing a reference sound played two times and a comparison sound played once, the participants will be asked to mark whether the comparison sound was the same as the reference sound or not. Additionally, each participant may mark whether or not they were familiar with each of the melodies presented to provide concrete evidence of whether familiarity with the given stimuli facilitates their perception or not. I propose playing each of the types of the stimuli to the participants before the experiment itself to ensure they understand the task clearly.

2.1.3 Stimuli

The stimuli should ideally consist of acoustically rich sounds since they are more readily and easily processed as compared to pure tones (Iino et al. 2018) as mentioned in chapter 1.5.3. Pure tones may be used, however, in such a case, extra attention should be paid to the age of the participants (Iino et al. 2018). If both the types of stimuli are used, i.e., complex tones and pure tones, it is important to keep them separate in the results to avoid potential bias of the data. As discussed, vocal melodies are more salient to humans than instrumental melodies (Weiss et al. 2012; 2015). This does not necessarily mean that one or the other should be excluded, however, the resulting data should, again, be kept separate if both vocal and instrumental melodies are used. Any melodies used in the stimuli should conform to the standards of Western music as it is this type of music that the majority of English and Czech native speakers are accustomed to. Western music melodies must not be mixed with melodies that follow the musical structures of different cultures as the participants may have difficulties processing the unfamiliar structures (Trainor and Trehub 1992; Trehub et al. 1999). A wide variety of the stimuli may provide further insights into the field of musical pitch perception, targeting specific complexities such as whether there is a difference in perception of stimuli containing both instrumental background and a leading singer as opposed to seemingly simpler stimuli in the form of a note-by-note melody. The following suggested stimuli are a combination of the stimuli used in both PROMS and MIQ.

I suggest creating at least four different sets of stimuli. Firstly, to test for participants' "coarse-grained" pitch perception abilities, two sets of stimuli may be used: 1) vocal Western melodies and 2) instrumental Western melodies. The comparison melody may then contain a mistake in the form of a changed interval (at least by one semi-tone but preferably larger). Such larger changes in a melody should be perceived by listeners via the "coarse-grained" pitch perception. To test for subjects' "fine-grained" pitch perception, reference melodies should be presented in the same way as when testing for the "coarse-grained" perception. The comparison melody, however, when different from the reference, should contain only a slightly mistuned tone. The degree of mistuning should vary (however, it should not be mistuned more than by a semi-tone and probably should not even closely approximate the distance of one semi-tone), to test the perception of both more evident changes, as well as very fine mistuning differences. To provide further variety, 3) chords, basic triads (consisting of the root, the major or minor third, and the fifth), as well as more complex ones, may be presented to the participants. The comparison chord may then contain a wrong note which will result in a dissonant sound, or one of the notes will be slightly mistuned, potentially evoking in the participants the feeling that it is out of tune. Next, instead of the pure tones used in PROMS, 4) single acoustically rich sounds in isolation may be presented to the participants to avoid the discussed effect of pure tones. Some comparison sounds should then be slightly mistuned from the reference. Additionally, 5) an extract from a song that includes both instruments and a leading singer may be presented to the participants. In the comparison sample, the singer or the band may then be slightly mistuned in relation to the other.

2.1.4 Participants

The participants should be native speakers of Czech and English. In a perfect scenario, the participants should not speak any other languages to avoid the potential influence of L2, however, such participants would be extremely difficult to find. Importantly, though, no concessions should be made when it comes to the participants' experience with tonal languages.

Although most of the studies reviewed in this paper with relation to this topic focus on the effect that tonal language as L1 has on one's musical abilities (Pfordresher and Brown 2009; Bidelman et al. 2013; Choi 2021; Chen et al. 2016), it can be expected that even an elemental experience with a tonal language as L2 may have similar effects on one's musical abilities. The participants should therefore consist of native speakers of Czech and English, both men and women, who have no experience with tonal languages, i.e., they are not learning or fluently speaking any such language. Participants should be of all ages, including children. Based on self-report, participants should be divided into musicians and non-musicians according to whether, and if so then for how long they have been taking music lessons, and how often they come into contact with music (i.e., how many hours per week they listen to music, play an instrument, or sing). The resulting data need to be closely examined to spot where a potential difference between accordingly divided musicians and non-musicians occurs. Keeping track of the participants' sex and age is necessary in case any distinction based on these differences occurs. Last but not least, all participants should self-report whether or not they suffer from any hearing or speaking difficulties or disorders. Such participants may still be involved in the experiment; however, for obvious reasons they cannot be mixed with healthy subjects in the results.

2.2 Testing Musical Pitch Production

In this section, I will propose adequate methodology for testing musical pitch production. I will suggest adequate stimuli and procedure that may serve as a guide, as well as just an inspiration for any further research.

2.2.1 Procedure

Testing musical pitch production via an online experiment is questionable as the acoustic quality and intelligibility of the data collected are not ensured. I therefore suggest testing musical pitch production in person in a completely silent room with a quality microphone that ensures effortless intelligibility.

First, participants should fill in the same questionnaire as proposed for the testing of musical pitch perception in chapter 2.1.2. Then, they will be asked to listen to the reference stimuli and to imitate it vocally using the consonant-vowel construction "la" or "na". I propose using the first five or so stimuli as a test to ensure that participants understand their task and perhaps also to reduce the potential initial stress. Participants should be instructed to produce their vocal imitations clearly and loudly with a strong voice. The experimenter may even produce the first few test stimuli together with the participants to potentially evoke a feeling of trust towards the experimenter which may lead to further reduction of the stress commonly associated with singing in front of other people. Stress may lead to quiet shaky uncertain voice which negatively affects the ability to sing in tune. Additionally, the experimenter may leave the room after the test round so as not to unsettle the participants.

2.2.2 Stimuli

The stimuli may be a combination of the types of stimuli used in the experiment testing for pitch perception described in chapter 2.1.3. The minimal set of stimuli should consist of two groups. The first group should contain acoustically rich isolated tones. The second group should contain vocal or instrumental Western melodies. The melodies should not be very fast and long as this experiment does not focus on articulation or short-term memory but rather on the spectral quality of the production.

2.2.3 Participants

Last but not least, the participants of an experiment testing musical pitch production should meet the same criteria as the subjects tested for musical pitch perception discussed in the previous chapter (2.1.4).

Conclusion

This thesis aims to bring new insights into the field of music perception and production, specifically, how the two processes may be influenced by speakers' native language. Special focus is given to the effect of non-tonal language background which is a field much less examined than the field of tonal languages. Namely, based on the fact that English employs wider fundamental frequency (F0) contours in everyday speech than Czech, I hypothesise that English speakers have more refined pitch perception and pitch production abilities than Czech speakers whose native speech is often described as monotonous due to its flat F0 contours. In the first chapter, I review relevant literature and previous findings related to this topic, describing the development, perception, and the importance of F0 in the domains of speech and music, and provide evidence for the possibility of the language-to-music transfer even for non-tonal languages. Then, the cognitive processes underlying speech and music perception are discussed, and afterwards, the F0 contours of each of the two languages, English and Czech, are described in detail. Finally, individual factors that need to be controlled for when testing musical pitch perception and production are discussed. The second chapter is dedicated to an experiment proposal targeting both musical pitch perception and production. Initially, two already existing online experiments testing musical pitch perception are provided and commented on. Then, drawing inspiration from their methodology, I propose adequate methodology for creating new experiments testing both musical pitch perception and production. Since carrying out the experiments would exceed the extent of this thesis, using the proposed methodology to carry out the experiments and collect the data is highly encouraged. The data collected would provide unique insights into this specific field as such difference between Czech and English native speakers has, to the best of my knowledge, not yet been explored. Further research targeting possible differences between the musical abilities of native speakers of various non-tonal languages that differ only in the range of the F0 contours they employ may provide a better understanding of the possible language-to-music transfer discussed in this paper.

Bibliography

Akahane-Yamada, Reiko, Yoh'ichi Tohkura, Ann R. Bradlow, and David B. Pisoni. 1996. "Does Training in Speech Perception Modify Speech Production?" Paper presented at the *Proceeding of Fourth International Conference on Spoken Language Processing, ICSLP'96*, Philadelphia, PA, USA, October 1996, vol. 2, part 1 (3 October–6 October): 606–609. IEEE. https://doi.org/10.1109/ICSLP.1996.607434.

Alexander, Jennifer A., Patrick CM Wong, and Ann R. Bradlow. 2005. "Lexical Tone Perception in Musicians and Non-Musicians." Paper presented at *Interspeech*, Lisbon, Portugal, September 2005, (4 September–8 September): 397–400. https://doi.org/10.21437/Interspeech.2005-271.

Asaridou, Salomi S., and James M. McQueen. 2013. "Speech and Music Shape the Listening Brain: Evidence for Shared Domain-General Mechanisms." *Frontiers in Psychology* 4, no. 321 (June): 1–14. https://doi.org/10.3389/fpsyg.2013.00321.

Besson, Mireille, Julie Chobert, and Céline Marie. 2011. "Transfer of Training Between Music and Speech: Common Processing, Attention, and Memory." *Frontiers in Psychology* 2, no. 1–12. (May): 94. https://doi.org/10.3389/fpsyg.2011.00094.

Bidelman, Gavin M., Jackson T. Gandour, and Ananthanarayan Krishnan. 2011a. "Musicians and Tone-Language Speakers Share Enhanced Brainstem Encoding but Not Perceptual Benefits for Musical Pitch." *Brain and Cognition* 77, no. 1 (October): 1–10. https://doi.org/10.1016/j.bandc.2011.07.006.

Bidelman, Gavin M., Jackson T. Gandour, and Ananthanarayan Krishnan. 2011b. "Cross-Domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem." *Journal of Cognitive Neuroscience* 23, no. 2 (February): 425–434. https://doi.org/10.1162/jocn.2009.21362.

Bidelman, Gavin M., Stefanie Hutka, and Sylvain Moreno. 2013. "Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality Between the Domains of Language and Music." *PloS one* 8, no. 4 (April): e60676. https://doi.org/10.1371/journal.pone.0060676.

Casserly, Elizabeth D., and David B. Pisoni. 2010. "Speech Perception and Production." *Wiley Interdisciplinary Reviews: Cognitive Science* 1, no. 5 (July): 629–647. https://doi.org/10.1002/wcs.63.

Chamonikolasová, Jana. 2017. *Intonation in English and Czech Dialogues*, 2nd edition. Spisy Filozofické fakulty Masarykovy univerzity, vol. 365. Brno: Masarykova univerzita. ISBN 978-80-210-9119-1. http://hdl.handle.net/11222.digilib/137012.

Chang, Daniel, Nancy Hedberg, and Yue Wang. 2016. “Effects of Musical and Linguistic Experience on Categorization of Lexical and Melodic Tones.” *The Journal of the Acoustical Society of America* 139, no. 5 (May): 2432–2447. https://doi.org/10.1121/1.4947497.

Chen, Ao, Liquan Liu, and René Kager. 2016. “Cross-Domain Correlation in Pitch Perception, the Influence of Native Language.” *Language, Cognition and Neuroscience* 31, no. 6 (April): 751–760. https://doi.org/10.1080/23273798.2016.1156715

Chládková, Kateřina, and Nikola Paillereau. 2020. “The What and When of Universal Perception: A Review of Early Speech Sound Acquisition.” *Language Learning* 70, no. 4 (July): 1136–1182. https://doi.org/10.1111/lang.12422.

Choi, William. 2021. “Musicianship Influences Language Effect on Musical Pitch Perception.” *Frontiers in Psychology* 12, (October): 712753. https://doi.org/10.3389/fpsyg.2021.712753.

Cirelli, Laura K., and Sandra E. Trehub. 2020. “Familiar Songs Reduce Infant Distress.” *Developmental Psychology* 56, no. 5: 1–29. https://psycnet.apa.org/doi/10.1037/dev0000917.

Cooper, Angela, and Yue Wang. 2012. “The Influence of Linguistic and Musical Experience on Cantonese Word Learning.” *The Journal of the Acoustical Society of America* 131, no. 6 (June): 4756–4769. https://doi.org/10.1121/1.4714355.

Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics*, 6th ed. Oxford: Blackwell Publishing Ltd.

Deutsch, Diana. 2013. “Absolute Pitch.” *The Psychology of Music* 3 (January): 141–182. https://doi.org/10.13140/2.1.4752.7048.

Deutsch, Diana. 2002. “The Puzzle of Absolute Pitch.” *Current Directions in Psychological Science* 11, no. 6 (December): 200–204. https://doi.org/10.1111/1467-8721.00200.

Deutsch, Diana, Kevin Dooley, Trevor Henthorn, and Brian Head. 2009. “Absolute Pitch among Students in an American Music Conservatory: Association with Tone Language Fluency.” *The Journal of the Acoustical Society of America* 125, no. 4 (April): 2398–2403. https://doi.org/10.1121/1.3081389.

Deutsch, Diana, Trevor Henthorn, and Mark Dolson. 2004a. “Absolute Pitch, Speech, and Tone Language: Some Experiments and a Proposed Framework.” *Music Perception* 21, no. 3 (March): 339–356. https://doi.org/10.1525/mp.2004.21.3.339.

Deutsch, Diana, Trevor Henthorn, and Mark Dolson. 2004b. “Speech Patterns Heard Early in Life Influence Later Perception of the Tritone Paradox.” *Music Perception* 21, no. 3 (March): 357–372. https://doi.org/10.1525/mp.2004.21.3.357.

Deutsch, Diana, Trevor Henthorn, and Rachael Lapidis. 2011. “Illusory Transformation from Speech to Song.” *The Journal of the Acoustical Society of America* 129, no. 4 (April): 2245–2252. https://doi.org/10.1121/1.3562174.

Deutsch, Diana, Trevor Henthorn, Elizabeth Marvin, and HongShuai Xu. 2006. “Absolute Pitch among American and Chinese Conservatory Students: Prevalence Differences, and Evidence for a Speech-Related Critical Period.” *The Journal of the Acoustical Society of America* 119, no. 2 (February): 719–722. https://doi.org/10.1121/1.2151799.

Frazier, Lyn, Katy Carlson, and Charles Clifton. 2006. “Prosodic Phrasing is Central to Language Comprehension.” *Trends in Cognitive Sciences* 10, no. 6 (June): 244–249. https://doi.org/10.1016/j.tics.2006.04.002.

Honorof, Douglas N., and Douglas H. Whalen. 2005. “Perception of Pitch Location within a Speaker’s F0 Range.” *The Journal of the* Acoustical Society of America 117, no. 4 (April): 2193–2200. https://doi.org/10.1121/1.1841751.

Hutka, Stefanie, Gavin M. Bidelman, and Sylvain Moreno. 2015. “Pitch Expertise Is Not Created Equal: Cross-Domain Effects of Musicianship and Tone Language Experience on Neural and Behavioural Discrimination of Speech and Music.” *Neuropsychologia* 71 (March): 52–63. https://doi.org/10.1016/j.neuropsychologia.2015.03.019.

Iino, Hiroko, Katsuya Ohta, Keiko Hara, Miho Miyajima, Minoru Hara, Eisuke Matsushima, and Masato Matsuura. 2018. “Vowel-Speech Versus Pure-Tone Processing in Healthy Subjects.” *Neuroscience Research* 137 (December): 43–48. https://doi.org/10.1016/j.neures.2018.03.002.

Jekiel, Mateusz, and Kamil Malarski. 2021. “Musical Hearing and Musical Experience in Second Language English Vowel Acquisition.” *Journal of Speech, Language, and Hearing Research* 64, no. 5 (May): 1666–1682. https://doi.org/10.1044/2021_JSLHR-19-00253.

Koelsch, Stefan, Thomas C. Gunter, D. Yves V. Cramon, Stefan Zysset, Gabriele Lohmann, and Angela D. Friederici. 2002. “Bach Speaks: A Cortical “Language-Network” Serves the Processing of Music.” *Neuroimage* 17, no. 2 (October): 956–966. https://doi.org/10.1006/nimg.2002.1154.

Ladefoged, Peter, and Keith Johnson. 2014. *A Course in Phonetics*, 7th ed. Stamford: Cengage Learning.

Laštůvka, Jan. 2023. “Annotating English Intonation in ToBI with a Comparison with Czech.” Bachelor’s Thesis, Univerzita Palackého v Olomouci. https://theses.cz/id/4hmcrc/Bachelor_thesis_-_Jan_Lastuvka.pdf.

Levis, John M., and Anne Wichmann. 2015. “English Intonation – Form and Meaning.” In *The Handbook of English Pronunciation,* edited by Marnie Reed and John M. Levis. 139–155. Chichester: John Wiley & Sons Inc.

Levitin, Daniel J., and Susan E. Rogers. 2005. "Absolute Pitch: Perception, Coding, and Controversies." *Trends in Cognitive Sciences* 9, no. 1 (January): 26–33. https://doi.org/10.1016/j.tics.2004.11.007.

Liu, Jingxuan, Courtney B. Hilton, Elika Bergelson, and Samuel A. Mehr. 2023. "Language Experience Predicts Music Processing in a Half-Million Speakers of Fifty-Four Languages." *Current Biology* 33, no. 10 (May): 1916–1925. https://doi.org/10.1016/j.cub.2023.03.067.

Maess, Burkhard, Stefan Koelsch, Thomas C. Gunter, and Angela D. Friederici. 2001. "Musical Syntax Is Processed in Broca's Area: an MEG Study." *Nature Neuroscience* 4, no. 5 (May): 540–545. https://doi.org/10.1038/87502.

Mampe, Birgit, Angela D. Friederici, Anne Christophe, and Kathleen Wermke. 2009. "Newborns' Cry Melody Is Shaped by Their Native Language." *Current Biology* 19, no. 23 (December): 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064.

Mantell, James T., and Peter Q. Pfordresher. 2013. "Vocal Imitation of Song and Speech." *Cognition* 127, no. 2 (May): 177–202. https://doi.org/10.1016/j.cognition.2012.12.008.

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, no. 9 (September): 1453–1463. https://doi.org/10.1162/jocn.2007.19.9.1453.

May, Lillian, Judit Gervain, Manuel Carreiras, and Janet F. Werker. 2018. "The Specificity of the Neural Response to Speech at Birth." *Developmental Science* 21, no. 3 (May): e12564. https://doi.org/10.1111/desc.12564.

Montagni, Ilaria, and Andrea Peru. 2012. "Effects of Language and Music Learning on Pitch and Duration Perception: An Experimental Investigation." *International Journal of Psychology and Behavioral Sciences* 1, no. 1 (December): 33–40. https://doi.org/10.5923/j.ijpbs.20110101.05.

Musemap. 2024. "Profile of Music Perception Skills (PROMS)." Musemap. April 4, 2024. https://musemap.org/resources/proms.

Ngo, Mary Kim, Kim-Phuong L. Vu, and Thomas Z. Strybel. 2016. "Effects of Music and Tonal Language Experience on Relative Pitch Performance." *The American Journal of Psychology* 129, no. 2 (Summer): 125–134. https://doi.org/10.5406/amerjpsyc.129.2.0125.

Ohala, John J. 1984. "An Ethological Perspective on Common Cross-Language Utilization of F_0 of Voice." *Phonetica* 41, no. 1 (January): 1–16. https://doi.org/10.1159/000261706.

Palková, Zdena. 1994. *Fonetika a fonologie češtiny*. Praha: Karolinum.

Patel, Aniruddh D. 2011. "Why Would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis." *Frontiers in Psychology* 2, no. 142 (June): 1–14. https://doi.org/10.3389/fpsyg.2011.00142.

Peretz, Isabelle, Sébastien Nguyen, and Stéphanie Cummings. 2011. "Tone Language Fluency Impairs Pitch Discrimination." *Frontiers in Psychology* 2, no. 145 (July): 145. https://doi.org/10.3389/fpsyg.2011.00145.

Peter, Beate, Bronsyn Foster, Heather Haas, Kyle Middleton, and Kiersten McKibben. 2014. "Direct and Octave-Shifted Pitch Matching During Nonword Imitations in Men, Women, and Children." *Journal of Voice* 29, no. 2 (November): P260.e21–260.e30. https://doi.org/10.1016/j.jvoice.2014.06.011.

Pfordresher, Peter Q., and Steven Brown. 2009. "Enhanced Production and Perception of Musical Pitch in Tone Language Speakers." *Attention, Perception, & Psychophysics* 71, no. 6 (August): 1385–1398. https://doi.org/10.3758/APP.71.6.1385.

Pfordresher, Peter, and James Mantell. 2009. "Singing as a Form of Vocal Imitation-Mechanisms and Deficits." Paper presented at the *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music, ESCOM 2009*, Jyväskylä, Finland. 425–430. http://urn.fi/URN:NBN:fi:jyu-2009411309.

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

Pulvermüller, Friedemann, and Yury Shtyrov. 2006. "Language Outside the Focus of Attention: The Mismatch Negativity as a Tool for Studying Higher Cognitive Processes." *Progress in Neurobiology* 79, no. 1 (June): 49–71. https://doi.org/10.1016/j.pneurobio.2006.04.004.

Sergeant, Desmond. 1969. "Experimental Investigation of Absolute Pitch." *Journal of Research in Music Education* 17, no. 1 (Spring): 135–143. https://doi.org/10.2307/3344200.

Shattuck-Hufnagel, Stefanie, and Alice E. Turk. 1996. "A Prosody Tutorial for Investigators of Auditory Sentence Processing." *Journal of Psycholinguistic Research* 25, no. 2: 193–247. https://doi.org/10.1007/BF01708572.

Skarnitzl, Radek, and Hana Hledíková. 2022. "Prosodic Phrasing of Good Speakers in English and Czech." *Frontiers in Psychology* 13 (March): 857647. https://doi.org/10.3389/fpsyg.2022.857647.

Slevc, Robert L., and Akira Miyake. 2006. "Individual Differences in Second-Language Proficiency: Does Musical Ability Matter?" *Psychological Science* 17, no. 8 (August): 675–681. https://journals.sagepub.com/doi/10.1111/j.1467-9280.2006.01765.x.

Slevc, Robert L., Jason C. Rosenberg, and Aniruddh D. Patel. 2009. “Making Psycholinguistics Musical: Self-Paced Reading Time Evidence for Shared Processing of Linguistic and Musical Syntax.” *Psychonomic Bulletin & Review* 16, no. 2 (April): 374–381. https://doi.org/10.3758/16.2.374.

Stevens, Catherine J., Peter E. Keller, and Michael D. Tyler. 2013. “Tonal Language Background and Detecting Pitch Contour in Spoken and Musical Items.” *Psychology of Music* 41, no. 1 (January): 59–74. https://doi.org/10.1177/0305735611415749.

Swaminathan, Swathi, Haley E. Kragness, and E. Glenn Schellenberg. 2021. “The Musical Ear Test: Norms and Correlates from a Large Sample of Canadian Undergraduates.” *Behaviour Research Methods* 53, (March): 2007–2024. https://doi.org/10.3758/s13428-020-01528-8.

Takeuchi, Annie H., and Stewart H. Hulse. 1993. “Absolute Pitch.” *Psychological Bulletin* 113, no. 2: 345–361. https://doi.org/10.1037/0033-2909.113.2.345.

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 (April): 1–7. https://doi.org/10.1016/j.lindif.2018.04.003.

The Music Lab. 2024. “Test Your Musical IQ.” The Musical Lab. N.d. https://www.themusiclab.org/quizzes/miq.

Trainor, Laurel J., and Sandra E. Trehub. 1992. “A Comparison of Infants' and Adults' Sensitivity to Western Musical Structure.” *Journal of Experimental Psychology: Human Perception and Performance* 18, no. 2 (May): 394–402. https://doi.org/10.1037/0096-1523.18.2.394.

Trehub, Sandra E., E. Glenn Schellenberg, and Stuart B. Kamenetsky. 1999. “Infants' and Adults' Perception of Scale Structure.” *Journal of Experimental Psychology: Human Perception and Performance* 25, no. 4: 965–975. https://doi.org/10.1037/0096-1523.25.4.965.

Trehub, Sandra E., Anna M. Unyk, Stuart B. Kamenetsky, David S. Hill, Laurel J. Trainor, Joanna L. Henderson, and Myra Saraza. 1997. “Mothers' and Fathers' Singing to Infants.” *Developmental Psychology* 33, no. 3: 500–507. https://doi.org/10.1037/0012-1649.33.3.500.

Volín, Jan, Damien Galeone, and Wesley Johnson. 2017. “Pitch Range of Intonation Contours in English Czech.” *AUC PHILOLOGICA* 2017, no. 3 (September): 55–64. https://doi.org/10.14712/24646830.2017.33.

Volín, Jan, Kristýna Poesová, and Lenka Weingartová. 2015. “Speech Melody Properties in English, Czech and Czech English: Reference and Interference.” *Research in Language* 13, no. 1 (March): 107–123. https://doi.org/10.1515/rela-2015-0018.

Vonzová, Veronika. 2023. "Melodic Contours in Political Debates: Comparison of British and American English." Master's Thesis, Univerzita Karlova. https://dspace.cuni.cz/bitstream/handle/20.500.11956/179424/120438903.pdf?sequence=1.

Warren, Jason. 2008. "How Does the Brain Process Music?" *Clinical Medicine* 8, no. 1 (February): 32–36. https://doi.org/10.7861/clinmedicine.8-1-32.

Wayland, Ratree, Elizabeth Herrera, and Edith Kaan. 2010. "Effects of Musical Experience and Training on Pitch Contour Perception." *Journal of Phonetics* 38, no. 4 (October): 654–662. https://doi.org/10.1016/j.wocn.2010.10.001.

Weiss, Michael W., E. Glenn Schellenberg, Sandra E. Trehub, and Emily J. Dawber. 2015. "Enhanced Processing of Vocal Melodies in Childhood." *Developmental Psychology* 51, no. 3: 370–377. https://doi.org/10.1037/a0038784.

Weiss, Michael W., Sandra E. Trehub, and E. Glenn Schellenberg. 2012. "Something in the Way She Sings: Enhanced Memory for Vocal Melodies." *Psychological Science* 23, no. 10 (August): 1074–1078. https://doi.org/10.1177/0956797612442552.

Weiss, Michael W., Sandra E. Trehub, E. Glenn Schellenberg, and Peter Habashi. 2016. "Pupils Dilate for Vocal or Familiar Music." *Journal of Experimental Psychology: Human Perception and Performance* 42, no. 8: 1061–1065. https://doi.org/10.1037/xhp0000226.

Wells, John C. 2006. *English Intonation: An Introduction*. New York: Cambridge University Press.

Wells, John C. 2014. *Sounds Interesting: Observations on English and General Phonetics*. Cambridge: Cambridge University Press.

Wiener, Seth, and Evan D. Bradley. 2023. "Harnessing the Musician Advantage: Short-Term Musical Training Affects Non-Native Cue Weighting of Linguistic Pitch." *Language Teaching Research* 27, no. 4 (July): 1016–1031. https://doi.org/10.1177/1362168820971791.

Wong, Patrick C. M., Valter Ciocca, Alice H. D. Chan, Louisa Y. Y. Ha, Li-Hai Tan, and Isabelle Peretz. 2012. "Effects of Culture on Musical Pitch Perception." *PloS one* 7, no. 4 (April): e33424. https://doi.org/10.1371/journal.pone.0033424.

Zatorre, Robert J., and Shari R. Baum. 2012. "Musical Melody and Speech Intonation: Singing a Different Tune." *PLoS Biol* 10, no. 7 (July): e1001372. https://doi.org/10.1371/journal.pbio.1001372.

Zhang, Kaile, Xiao Wang, and Gang Peng. 2017. "Normalization of Lexical Tones and Nonlinguistic Pitch Contours: Implications for Speech-specific Processing Mechanism." *The Journal of the Acoustical Society of America* 141, no. 1 (January): 38–49. https://doi.org/10.1121/1.4973414.

Zhang, Linjun, Songcheng Xie, Yu Li, Hua Shu, and Yang Zhang. 2020. "Perception of Musical Melody and Rhythm as Influenced by Native Language Experience." *The Journal of the Acoustical Society of America* 147, no. 5 (May): EL385–EL390. https://doi.org/10.1121/10.0001179.

Annotation

Author: Vojtěch Kopecký
Field of study: English Philology
Title: Does Non-Tonal L1 Background Affect Musical Pitch Skills? Musical Abilities of Czech and English Native Speakers
Type of paper: Bachelor thesis
Faculty and Department: Faculty of Arts, Department of English and American Studies
Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D.
Number of pages: 41
Number of characters: 83,830

Description:
This thesis aims to explore the linguistic field of language-to-music transfer. It argues for the potential influence of native language background on one's musical pitch perception and production. Moreover, it discusses the potential difference in musical skills of native speakers of Czech and English, two non-tonal languages that differ notably in the width of the F0 contours they employ. As a result, research methodology for testing musical pitch perception and production is proposed for any potential future experiments.

Keywords:
Musical skills, pitch perception, pitch production, L1 background, non-tonal languages, speech, music, Czech, English, F0 contours

Anotace

Autor: Vojtěch Kopecký

Studijní obor: Anglická filologie

Název: Má netónový mateřský jazyk vliv na intonační schopnosti v hudbě? Hudební schopnosti českých a anglických rodilých mluvčích

Typ práce: bakalářská práce

Fakulta a katedra: Filozofická fakulta, Katedra anglistiky a amerikanistiky

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

Počet stran: 41

Počet znaků: 83.830

Charakteristika:

Tato práce se zabývá vlivem mateřského jazyka na hudební schopnosti. Detailně tento vliv porovnává mezi češtinou a angličtinou, dvěma netónovými jazyky, které se mezi sebou výrazně liší šíří intonačního rozpětí používaného v běžné mluvě. Na základě přehledu relevantní literatury je následně navržena metodologie experimentu zaměřujícího se na percepční a intonační schopnosti českých a anglických rodilých mluvčích.

Klíčová slova:

Intonační schopnosti, percepce hudebního tónu, produkce hudebního tónu, mateřský jazyk, netónové jazyky, hudba, řeč, čeština, angličtina, základní frekvence, tónové kontury