JSLHR

Research Article

Musical Hearing and Musical Experience in Second Language English Vowel Acquisition

Mateusz Jekiel[a] and Kamil Malarski[a]

Purpose: Former studies suggested that music perception can help produce certain accentual features in the first and second language (L2), such as intonational contours. What was missing in many of these studies was the identification of the exact relationship between specific music perception skills and the production of different accentual features in a foreign language. Our aim was to verify whether empirically tested musical hearing skills can be related to the acquisition of English vowels by learners of English as an L2 before and after a formal accent training course.

Method: Fifty adult Polish speakers of L2 English were tested before and after a two-semester accent training in order to observe the effect of musical hearing on the acquisition of English vowels. Their L2 English vowel formant contours produced in consonant–vowel–consonant context were compared with the target General British vowels produced by their pronunciation teachers. We juxtaposed these results with their musical hearing test scores and self-reported musical experience to observe a possible relationship between successful L2 vowel acquisition and musical aptitude.

Results: Preexisting rhythmic memory was reported as a significant predictor before training, while musical experience was reported as a significant factor in the production of more native-like L2 vowels after training. We also observed that not all vowels were equally acquired or affected by musical hearing or musical experience. The strongest estimate we observed was the closeness to model before training, suggesting that learners who already managed to acquire some features of a native-like accent were also more successful after training.

Conclusions: Our results are revealing in two aspects. First, the learners' former proficiency in L2 pronunciation is the most robust predictor in acquiring a native-like accent. Second, there is a potential relationship between rhythmic memory and L2 vowel acquisition before training, as well as years of musical experience after training, suggesting that specific musical skills and music practice can be an asset in learning a foreign language accent.

The assumed deep-binding interconnectivity between musical talent and the ease of learning foreign languages has been anecdotally repeated for a long time, yet too often without clear references to the relevant research. Indeed, there are many similarities between language and music, and both have been thoroughly studied over the years. Studies in music and language evolution argue that humans developed musicality in similar ways they developed their language capacities (Brown, 2001; Mithen, 2005; see also Baudeat, 2017). Another large portion of research focused on the similarities of how music and language are processed on the neural level (Brown et al., 2006; Chobert & Besson, 2013; Fadiga et al., 2009; Kunert et al., 2015). Considering these commonalities, language researchers and trainers have looked into the possible effects that music perception and production might have on the acquisition of various language structures (Pastuszek-Lipińska, 2008; Strait et al., 2012), as well as on its role in language teaching (Fonseca-Mora et al., 2011; Franklin et al., 2008; Picavet et al., 2012). Following this line of research, the current study is interested whether preexisting musical skills and musical experience can predict subsequent learning outcomes in the acquisition of second language (L2) pronunciation in a formal learning environment.

Despite the large body of research concerning the relationship between musical hearing and L2 learning, there is still considerable room for exploration due to the limitations of previous studies. First, many prior experiments relied on self-reported language proficiency (Roncaglia-

[a]Faculty of English, Adam Mickiewicz University, Poznań, Poland
Correspondence to Mateusz Jekiel: mjekiel@amu.edu.pl

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee
Received September 29, 2019
Revision received December 23, 2019
Accepted January 4, 2021
https://doi.org/10.1044/2021_JSLHR-19-00253

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Denissen et al., 2016), limited the notion of L2 proficiency to grammar and vocabulary (e.g., Franklin et al., 2008: musical training – verbal memory) or resorted to perceptual evaluation of L2 pronunciation (Milovanov et al., 2010). As far as music skills are concerned, many studies divided participants into musicians and nonmusicians based on their music education or assessed their musical aptitude via formal music education tests, which relied on both production and perception (Ott et al., 2011).

The primary aim of our research is to study the relationship between different aspects of musical hearing and the acquisition of L2 vowel inventory by Polish advanced learners of English. We conducted three musical hearing tests assessing pitch perception, melodic memory, and rhythmic memory, while music education and musical experience were treated independently. In order to assess segmental features of L2 pronunciation, we performed vowel measurements of our participants before and after a two-semester accent training course at a university level and compared their results with the model vowels of their pronunciation teachers. From all the second-language accent features, we have selected vowels because they can be reliably measured acoustically, while their signal properties are most similar to the music sequences comprising different tonalities. To summarize, the twofold goal of this study is to determine whether musical skills and musical experience can predict successful acquisition of L2 vowels during an accent training course, as well as to verify if these preexisting musical factors can be positively associated with L2 pronunciation prior to formal training.

Music, Language, and the Brain

Psycholinguists and cognitive neuroscientists (e.g., Chobert & Besson, 2013; Patel, 2008) emphasize that music and speech share similarities on various levels: (a) They are both specific to human beings and exist across all known cultures; (b) both are complex auditory signals that have tone, melody, and rhythm; (c) both can be described in terms of frequency, duration, intensity, and timbre; (d) both have several degrees of organization (morphology, phonology, semantics, syntax, and pragmatics in language; rhythm, melody, and harmony in music); (e) both systems require attention, memory, and sensorimotor coordination for accurate perception and production; and (f) both share neural resources for processing prosody, syntax, and semantics.

Neuroimaging and neuroscientific studies comparing the brain responses of musicians and nonmusicians when listening to musical and linguistic stimuli found functional and structural differences between both groups, suggesting that the properties of the transverse temporal gyri in the brain can be associated with musical aptitude (e.g., Schneider et al., 2002; see Barrett et al., 2013). Studies using functional magnetic resonance imaging and positron emission tomography also show an overlap in specific brain regions responsible for processing of both linguistic and musical structures, predominantly the Broca's area (e.g., Brown et al., 2006; Fadiga et al., 2009; Kunert et al., 2015), as well as greater brain plasticity and higher functional restructuring of the brain across people who had musical training at an earlier age (Gaser & Schlaug, 2003; Hyde et al., 2009); it has been shown that even short-termed musical training can be conducive to such neural changes (Lappe et al., 2008). While the abovementioned studies reveal an interesting relationship between musical skills and linguistic skills, Schellenberg (2004) and Patel (2012) point out that the majority of behavioral and neural data in favor of this hypothesis are not derived from experimental studies that would allow us to draw strong causal inferences. Importantly, however, a longitudinal study was conducted by Moreno et al. (2009) to directly test for causality (i.e., causal effects of music vs. painting training on different aspects of speech perception). Using both behavioral and electrophysiological methods, their results revealed that musical training causally impacts on linguistic abilities of children (i.e., positive transfers from musical training to speech processing). However, it still remains unclear which specific aspects of musical hearing potentially influence phonemic processing skills dependent on other acoustic properties. For example, if the speaker's timbre, an auditory attribute that relies on a different acoustic dimension than pitch, can be affected by musical training focused on pitch perception, it would suggest that both domains "share more abstract cognitive processes involved in sound categorization, possibly involving common cortical mechanisms" (Patel, 2012, p. 29).

The Role of Vowels in Language and Music

The sound systems in language and music are multimodal, complex, and therefore difficult to describe. Consonants can display secondary articulations; while producing vowels, in turn, speakers of different languages may make use of different tones that can change the meaning of an utterance. In music, timbre is also important, which becomes apparent when the same musical piece is transcribed to be played on a different instrument. In sung pieces, however, language and music are integrated: The singing is treated like an instrument, capable of operating on a particular timbre and pitch. It has been argued before that, at an earlier stage of language development, the modulations in timbre and pitch were more common (Nikolsky, 2015), and similar connections between vowel production and musical pitch are supported in experimental research on the processing of vowels and melody (Kolinsky et al., 2009; Russo et al., 2019). It has also been concluded that interacting neural networks are responsible for processing differences in phonemes and pitch in singing, treating them as integrated units (Lidji et al., 2010).

Vowels can be described as more music-like than consonants. First, Fenk-Oczlon (2017) emphasizes the importance of vowels for both language and music, pointing out their role in generating sonority in syllables and prosody in speech and singing. Similarly to musical notes, vowels have timbre, pitch, intensity, and duration. Moreover, the number of vowels often corresponds to the number of pitches in musical scales across cultures: Just like most languages have a vowel inventory comprising five vowels (Maddieson, 2013), pentatonic scales (i.e., musical scales with five notes per octave) are also considered to be the most common among traditional forms of music (Trehub et al., 1999). Finally, the musical quality of vowels plays a significant role in early language acquisition, especially in the use of pitch and timbre in early communication for pragmatic functions (Masataka, 2007). Due to these commonalities in the acoustic and auditory nature of vowels and music, we have selected this phonemic category for our analyses.

The common procedure in linguistics has been to present vowels in a two-dimensional space, most commonly on vowel charts, to approximate the position of the tongue during their production. Both the height/openness (F1) and frontness/backness (F2) of vowels are expressed in Hertz (Hz). In our analysis, we will focus only on the first (F1) and the second (F2) vowel formant characteristics because we are only interested in how high/low and front/back they are produced in our participants. Although it is possible for a person to pronounce a given vowel on different fundamental frequencies, the human brain must have a way to decode the F1 and F2 frequencies to hear the given vowel quality (see, e.g., Ikeda et al., 2014; Tankus et al., 2012).

Perception and Production

The relationship between perception and production in L2 pronunciation has been widely investigated over the years, frequently revolving around the notion of causal relationship between these two domains. The Speech Learning Model (Flege, 1995, 2007) postulates that the accuracy of production of nonnative sounds is correlated with their perception. This assumes that improved performance in perception is required for improved performance in production. Several studies have both confirmed and rejected this hypothesis (see Isbell, 2016). Another approach to study the relation between perception and production is correlational—recent research revealed moderate correlation between L2 perception and L2 production of segmental contrasts (e.g., Casillas, 2015); however, some studies found no correlation between these domains (cf. Zhang et al., 2016).

Latest findings show that production can shape perception, which means a language learner has to be constantly exposed to the spoken real-life target language structures; otherwise, production may disrupt the perceptual learning skills (Baese-Berk, 2019). An individual's first language (L1) can also influence their auditory mechanisms (e.g., phonetic properties of L1 speech can predict the performance in perception of L2 vowels across speakers; Kartushina & Frauenfelder, 2014). Moreover, speakers of different language can also process music and music-like sounds differently, for example, Chinese speakers demonstrate superior pitch processing (Bidelman et al., 2011), while Finnish speakers display more precise discrimination of duration (Dawson et al., 2017). Such differences are associated with the structure of an L1 and its cognitive representation in the speaker.

Musical Hearing and L2 Acquisition

Successful acquisition of L2 pronunciation relies on a number of linguistic and extralinguistic factors. The former include similarities and differences between L1 and L2 phonetic inventories, while the latter involve such aspects as the onset age of exposure to L2, motivation, learning strategies, working memory, and musical experience. According to Chobert and Besson (2013, p. 923), musical hearing "positively influences several aspects of speech processing, from auditory perception to speech production." In this study, we are interested whether musical hearing is associated with the production of L2 vowels before and after formal accent training.

According to Dolman and Spring (2014), musical hearing can be defined as an untaught, natural musical ability, and should be regarded as separate from musical experience, which includes music lessons and musicianship. One traditional method of assessing musical hearing is the Seashore Measures Of Musical Talents (Seashore et al., 1960), a standardized test that divides musical hearing into separate talents (focusing on pitch, duration, rhythm, timbre and tonality, loudness, etc.). However, this type of assessment is very formal and used primarily at music schools, where the participants are potential candidates already familiar with the fundamentals of music theory and practice. Here, we use a series of musical hearing tests devised by Mandell (2009), which rely on music perception and can be attempted by participants without any musical knowledge or formal instruction.

Recent studies suggest that musical hearing and training can influence second language acquisition, especially the acquisition of L2 pronunciation, including segmental and suprasegmental vocalic discrimination (Chobert & Besson 2013). Musical aptitude seems to correlate with phonological processing ability among preschool children (e.g., Anvari et al., 2002), while musical training has been associated with better linguistic skills (Strait et al., 2012; Tallal & Gaab, 2006), preexisting musical abilities seem to be better predictors for the acquisition of language skills than formal training in music (Swaminathan & Schellenberg, 2019). In a study by Slevc and Miyake (2006), superior musical hearing correlated with more accurate L2 listening discrimination and production skills among Japanese adult learners of English, while Milovanov et al. (2010) observed a relationship between musical skills and pronunciation skills among Finnish adult learners of English producing challenging English phonemes in a speech shadowing task. These findings motivated the current research, which is aimed at establishing whether musical hearing skills and former musical experience can be a valid predictor for successful acquisition of L2 pronunciation among Polish adult learners of English.

Polish Learners of English

The Polish vowel system is much less complicated than the one we find in English. Also, dialectal variation in vowels is marginal, as opposed to the large differentiation in English.

Polish has six vowels (/a/, /ɛ/, /i/, /ɨ/, /ɔ/, /u/) that do not differ significantly in length phonemically. Considering the rich vowel inventory in English, where vowels differ both in terms of quantity and quality, it is natural that Polish students of English usually find it difficult to master these contrasts. One problematic vowel contrast is the KIT–FLEECE distinction, in terms of both production and perception (Rojczyk & Porzuczek, 2012, p. 99; Sobkowiak, 2008), despite the fact that, in Polish, there are seemingly comparable vowel categories /i/ and /ɨ/. The TRAP vowel also poses many problems for Polish learners, even at advanced levels of language proficiency—the reason for this is the lack of a counterpart for this vowel in Polish, resulting in the assimilation to /ɛ/ or /a/ (Weckwerth, 2011).

To the best of our knowledge, there have only been a few studies investigating the role of music in the acquisition of L2 English phonology featuring participants with L1 Polish. A study by Pastuszek-Lipińska (2008) showed that Polish learners of English with formal music expertise (i.e., musical education and training) tend to outperform learners without any musical background in imitation and shadowing tasks. In this series of observations, musicians were better at discriminating foreign language sounds and produced fewer errors. A similar study by Zybert and Stępień (2009) utilized tests assessing music perception (Edwin Gordon's *Intermediate Measure of Music Audiation*) and production (imitating musical intervals), along with a speech perception and production test involving textbook audio recordings of a native speaker of English. The results for Polish secondary school learners of English with and without music education showed a correlation between perception and production of language and music, suggesting that both music perception and production can be good predictors in successful L2 learning. A fairly recent pilot study by Gralińska-Brawata and Rybińska (2017) investigated the production of word stress and its connection with musical abilities, where Polish advanced learners of English were asked to answer a questionnaire related to music, perform a music perception and production test, and read a set of commonly mispronounced words. The results suggest a possible relation between the correct word stress and superior musical abilities, although the authors pointed out the need for a larger sample. Since most of the abovementioned studies relied on imitation, shadowing, and repetition, the following study can be regarded as a relevant supplement to this discussion, as it incorporates visual stimuli and relies on learners' phonemic awareness developed throughout an intensive pronunciation training course.

This Study

The objective of the study is to investigate whether specific aspects of musical hearing and musical experience can predict successful acquisition of L2 vowels by Polish learners of English before and after formal accent training. The influence of musical hearing and musical training on the development of language skills, especially in the contexts of language acquisition and pedagogy, has recently received significant attention. The studies reported above, however, do not answer many of the questions we were interested in investigating. First, shadowing or imitation tasks do not explain the entire process in which L2 sounds are acquired, and in-class imitation is often dissimilar from the actual long-term formation of the L2 vowel inventory. For this reason, we conducted recording sessions before and after training, which relied on reading aloud from a screen a list of words without providing an immediate pronunciation model, which should offer a better insight into the acquisition of L2 pronunciation. Second, in order to sufficiently account for the musical hearing of our participants, we performed three different musical hearing tests assessing different aspects of musical hearing. We were interested in whether specific aspects of musical hearing, such as pitch perception, melodic memory, and rhythmic memory, are translatable into phonetic and phonological skills in our participants. Finally, we also included music education and musical experience in our analyses as separate factors that can relate to L2 vowel acquisition. To summarize, we formulate the following research questions:

1. Is there an observable progress in participants' L2 vowel production after training? Does the participants' distribution of vowel formants become more similar to their pronunciation teachers' vowel formants after training?
2. Are participants with better musical skills better at learning to pronounce L2 vowels?
3. Does musical experience predict how well participants learn to pronounce L2 vowels?

Method

Participants

Our subjects were 50 native speakers of Polish (42 women, eight men). Their age varied between 19 and 21 years ($M = 20.14$, $SD = 0.40$). At the time of the study, they were enrolled in Year 1 of a 3-year bachelor's program (1BA) in English studies, which included modules in linguistics, literature, culture, and EFL (English as a Foreign Language). All participants were recruited from four different groups that followed the same curriculum. At the time of the recording, their level of English was between B2 and C1 within the Common European Framework of Reference framework (a guide for categorizing European learners of foreign languages in terms of their achievement; Council of Europe, 2011). It was a homogenous group in this respect as they had performed very similarly in their secondary school final exams in written and oral English (Polish *matura* exams). To confirm their language proficiency, we also conducted LexTALE (Lexical Test for Advanced Learners of English) by Lemhöfer and Broersma (2012), commonly used to study participants with an advanced level of L2 English in an experimental setting. The results confirmed that the group was uniform and fairly advanced ($M = 74.48\%$, $SD = 8.93$). Despite their advanced command of written and spoken English, none of our participants had obtained substantial or regular pronunciation instruction at an earlier stage, which is not an exception, since pronunciation training in the EFL classroom is not systematically taught in primary or secondary education (Derwing & Munro, 2015). The first recording session was scheduled for the first week of their university education in order to factor out the influence of the accent training they were to receive during the academic year. Thus, so far, they had learnt their EFL pronunciation solely through naive exposure. None of our subjects had any medically documented speech or hearing impairments. They took part in the experiment in exchange for extra course credit. All participants were volunteers and were not financially remunerated for their participation. Informed written consent was obtained from all participants prior to their inclusion in the study.

Accent Training

All participants were assigned to a two-semester accent training course and attended the classes twice a week during the first year of their studies (a total of 90 hr of class work). The course comprised segmental phonetics (i. e., vowels and consonants of English) and suprasegmental phonetics (i.e., syllable stress, sentence stress, intonation, and rhythm), and its primary goal was to allow the participants to develop a consistent, native-like pronunciation based on the General British (henceforth GB) accent (i.e., the accent commonly used in education and usually associated with the South of England[1]). The course syllabus included such topics as English vowels (monophthongs and diphthongs) and consonants (fortis–lenis distinction and its effect on the preceding vowel), connected speech processes (e.g., assimilation, elision, linking), word stress, sentence stress, and weak forms.

Four different pronunciation teachers conducted the classes. The teachers were all female native speakers of Polish with native-like GB accents and were experienced language trainers with knowledge in phonetics and English language teaching. As the teachers shared the same L1 with the learners, they could rely on their individual experience in mastering English pronunciation and help the learners to avoid errors considered as typical for Polish learners of English (e.g., spelling pronunciation, difference in the tongue position, inconsistent vowel duration or lack of vowel reduction; see Sobkowiak, 2008). We recorded the teachers using the same procedure as for our participants in order to compare their vowel formant contours. The formant frequencies for their L2 vowels were similar to the model GB values found in Cruttenden (2014, p. 104): We observed a strong correlation between the teacher's L2 F1/F2 vowel formants and the model formant values (r = .99, df = 19, p < .001). Phonetic instruction used in the accent training course was holistic (i.e., instead of teaching the sounds in isolation, coarticulation, connected speech processes, and stress were incorporated throughout the course). In the classroom, the primary teaching methods involved exposure, phonetic transcription, production via drills and reading exercises, improvisation, and prepared speeches. Outside the classroom, all participants had to deliver monthly recordings based on the practiced reading material. Assessment relied on mid-semester and end-semester exams, which included a word list, a reading text, and spontaneous speech. The resulting accent was to be native-like and devoid of marked L1 features. After the training, all participants were expected to acquire native-like GB pronunciation through successful recognition and production of phonemic contrasts and native-like allophony.

English Phonetics and Phonology Course

The accent training course is also supplemented by an obligatory two-semester course in English phonetics and phonology, which provides a phonological description of the English sound system and raises phonological awareness. The primary aim of the course is to introduce students to the field of descriptive linguistics and help them as future educators in correcting errors made by Polish learners of English. The course was taught once a week for a total of 45 hr of class work, and the syllabus included the following topics: articulatory phonetics, acoustic phonetics, phonemic and phonetic transcription, English and Polish phonemes and allophones, connected speech processes, and prosody. Student assessment was based on weekly online quizzes and mid- and end-semester tests.

Recordings

Two recording sessions before and after training took place in a recording studio in a sound-treated booth. A single session, including the interview and musical hearing tests, took roughly 30 min for each participant. Each recording session was structured as a sociolinguistic interview and started with a casual conversation to minimize stress and help participants adapt to the experimental setting. Afterward, the main part of the recording session started, which comprised a series of word lists in English to elicit all GB monophthongs in the most formal and controlled context. The word list featured vowels in various consonant–vowel–consonant frames (b_d, b_t, d_d, d_t, g_d, h_d, and h_t) with five repetitions per word. The obtained vowels were 10 monophthongs: DRESS, KIT, FOOT, LOT, STRUT, TRAP, FLEECE, GOOSE, THOUGHT, and START. We obtained a total of 50 tokens in the h_d frame for this study for comparability with similar studies on vowel production, which often rely on this consonantal context. The word list was divided into 10 sections, each beginning with similarly phrased instructions: "These words rhyme with ___ or ___," (e.g., "These words rhyme with *feed* or *feet*"). The instructions were visualized on the screen and read by the investigator. Next, the participants were presented with the tokens belonging to that category on the monitor screen, always one

[1]At the beginning of their studies, students can choose to attend General British or General American accent training.

word per screen. The words were displayed in a large sans-serif font, in white letters against a black background. Each word was presented for 2,000 ms, with a 300-ms prerecording delay between each token. During that time frame, participants had to read aloud the token word that appeared on the screen. The stimulus was prepared and presented with SpeechRecorder software (Draxler & Jänsch, 2019), operated from a MacBook Pro. To capture the voice signal, we used an MXL 770 microphone, connected to a Roland Duo Capture EX audio interface. We were able to communicate with the participants sitting in the booth through AKG K240 MKII studio headphones and a PreSonus talkback monitor station. The speech samples were recorded at a mono 44.1-kHz frequency and 16-bit resolution, and were saved as separate wave files.

Musical Hearing Tests

After the first recording session but before accent training, each participant participated in three musical hearing tests designed by Mandell (2009), which were tested on over 11,000 subjects. Each test focuses on one musical hearing skill: pitch perception, melodic memory, and rhythmic memory, features that are regarded as valid indicators of musical aptitude and that are similarly tested in other musical hearing tests (e.g., Wallentin et al., 2010). The tests were run on a laptop connected to studio headphones.

The pitch perception test (Adaptive Pitch Test[2]) is designed to measure pitch perception abilities by playing a series of two short tones and asking if the second tone is higher or lower than the first. The participants used the UP and DOWN arrows on the keyboard to choose if the second tone was higher or lower. They could also replay the tones using the spacebar. The duration of the pitch perception test was approximately 1 min. Next, in the melodic memory test (Tonedeaf Test), each participant had to determine whether 36 pairs of short instrumental melodies were the same or different from one another by pressing the corresponding button on the screen. No repetition was possible in this test. The test was designed to verify pitch perception and melodic memory, as well as identify neuroanatomical correlates of tone deafness and was used as a screening test for patients tested in Mandell et al. (2007). The previously reported results had shown that the test was relatively difficult for both clinical (Mandell et al., 2007) and nonclinical (Mandell, 2009) participants; therefore, it is expected that our participants will also score low in this test. It took about 5 min to complete. Finally, in the rhythmic memory test (Rhythm Test), each participant was asked to decide whether each pair of 25 short rhythmic instrumental patterns was the same or different from each other by pressing the corresponding button on the screen. The test was designed to assess the ability to distinguish subtle differences in rhythm. The duration of this test was circa 5 min.

[2]The names in brackets are the names of the tests developed by Mandell (2009) available online.

Musical Experience Questionnaire

After the first recording session, we asked our participants in a questionnaire whether they attended music school or have specific musical experience (i.e., singing as soloists, band members or choir members, and/or playing a musical instrument as soloists, band members, or members of an orchestra or other ensemble). We also asked to specify the years spent in music school and/or years spent practicing singing and/or playing a musical instrument. Participants who reported no formal musical training or experience and who reported to sing or play a musical instrument only occasionally were treated in the analysis as nonmusicians.

Vowel Measurements

The data were prepared for analysis in Praat software (Boersma & Weenink, 2019), and the measurements were obtained through FAVE-Extract (Rosenfelder et al., 2014) and Montreal Forced Aligner (McAuliffe et al., 2017) using the DARLA web interface (Reddy & Stanford, 2015) and the *Vowels* R package (Kendall & Thomas, 2010). Since our participants were taught by different pronunciation teachers, we normalized their vowel measurement results separately for each group, including their teachers. We selected the Fabricius et al. (2009) method, which is speaker-intrinsic but vowel-extrinsic and formant-extrinsic. This normalization technique calculates a speaker's vowel space on the basis of the closest and the most open vowel in GB, which are often the TRAP and FLEECE vowels, respectively; in the Polish vowel system, the closest vowel is /i/, whereas the most open is /a/. By means of normalized vowel charts with the vowels of our participants plotted, we were able to graphically present the differences between their L2 vowels and the pronunciation model. The more similar their L2 vowels were to the model GB vowels, the better their EFL pronunciation would be evaluated. In order to quantify for the spatial relations between the vowels, we calculated the Euclidean distances between our participants' L2 vowels and teacher's model vowels (comparing F1 and F2 dimensions), where x1 stands for teacher's F1 and x2 for participant's F1, while y1 stands for teacher's F2 and y2 for participant's F2.

$$\text{distance} < \text{-sqrt}\left((x1-x2)^\wedge 2+(y1-y2)^\wedge 2\right). \qquad (1)$$

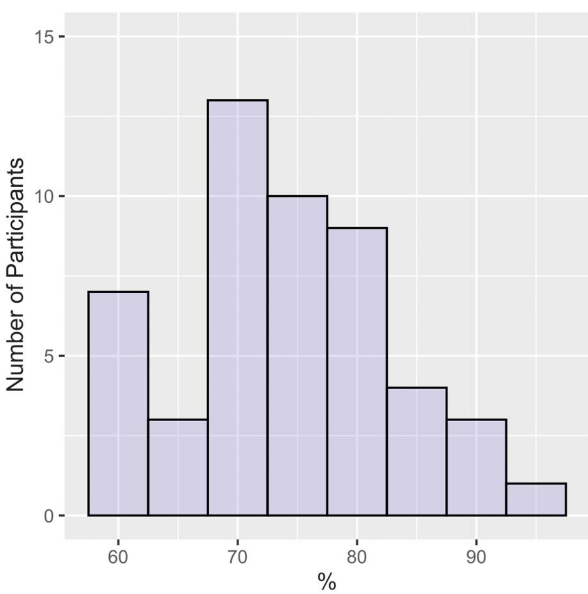
In the Results section, we are discussing the differences between the mean formant values of individual vowels produced by the participants and the corresponding mean vowel formants produced by the teachers. These vowel distance scores were then compared with the musical hearing test results, music education, years of musical experience, LexTALE results, and gender, using linear mixed-effects models.

Results

LexTALE

The LexTALE test results presented in Figure 1 confirmed that the subjects of the study formed a relatively

Figure 1. LexTALE results (%), the higher the better (M = 74.47, SD = 8.92).



uniform group in terms of their EFL proficiency. The horizontal axis is the test score (i.e., the percentage of correct responses). The mean result was 74.47 (min = 60, max = 97.50, Mdn = 73.75).

Musical Hearing Tests

The pitch perception test results presented in Figure 2 are expressed in Hertz (Hz) and indicate how reliably the participant could differentiate between two tones. The mean result was 16.05 Hz (min = 1 Hz, max = 60 Hz, Mdn = 10.02). The default grading for the test is fairly rigorous: The ability to reliably differentiate two tones less than 0.75 Hz apart indicates an exceptional ear (none of the participants achieved that score), while less than 1.5 Hz equals as fairly good (scored by 2 participants), less than 6 Hz as normal (13 participants), less than 12 Hz as low normal (13 participants) and above 16 Hz as below normal, possibly indicating a pitch perception deficit (18 participants).

The melodic memory test results in Figure 3 are expressed in percentages and indicate the percent of correctly identified tokens. According to the default scoring for this test, a result below 70% indicates low performance (scored by 27 participants), 70%–79% is normal (17 participants), 80%–90% is above normal (six participants), and above 90% is exceptional (none of the participants achieved that score). The mean result was 68% (min = 44.4%, max = 86.1%, Mdn = 69.4%).

The rhythmic memory test results presented in Figure 4 are expressed in percentages and indicate the percent of correctly identified tokens. Similarly to the previous test, a result below 70% indicates low performance, 70%–79% is normal, 80%–90% is above normal, and above 90% is exceptional. The mean result was 71.2% (min = 48%, max = 92%, Mdn = 72%).

Music Education and Musical Experience

Out of 50 participants, 18 confirmed to have had some musical experience. Four participants confirmed they had attended and completed the first degree of music school—two of them have played a musical instrument for 12 years

Figure 2. Pitch perception test results (Hz), the lower the better (M = 11.80, SD = 15.86).

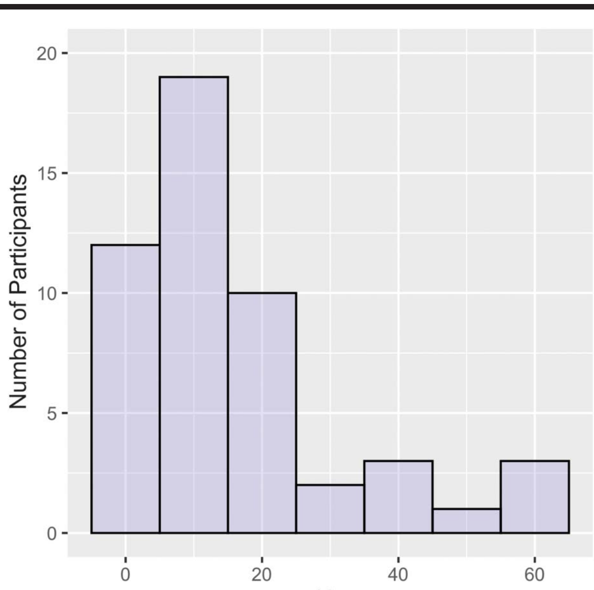


Figure 3. Melodic memory test results (%), the higher the better (M = 67.90, SD = 9.58).

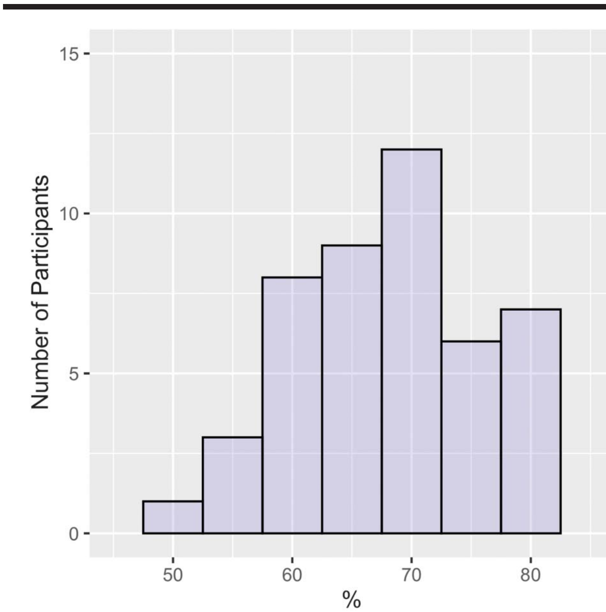


Figure 4. Rhythmic memory test results (%), the higher the better (M = 71.20, SD = 10.35).

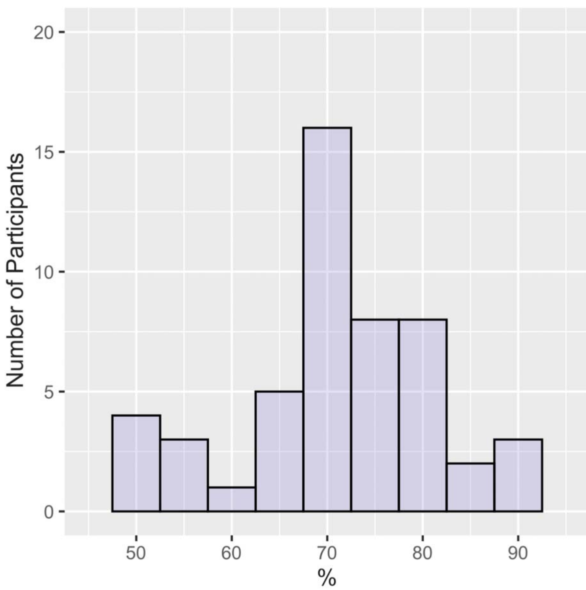
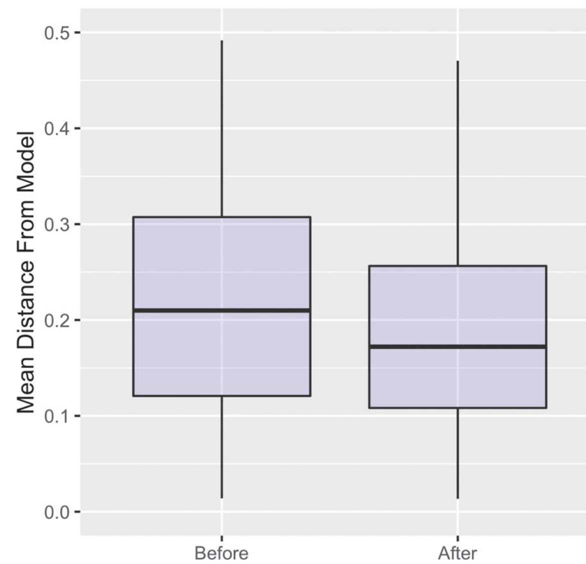


Figure 5. Distance from GB model before (M = 0.23, SD = 0.13) and after training (M = 0.20, SD = 0.12). The difference between the results is significant at $t(499) = -5.46$, $p < .001$.



and the other two have played for 6 years. Eight other participants reported to have played a musical instrument without formal music education—three of them have played for 9 years, four have played for 6 years, and one has played for 3 years. Thirteen participants confirmed singing either as soloists or band/choir members. Two of them were the same participants who had completed music school, while five were the same who reported to have played a musical instrument without formal training. Six other participants have sung for 12 years but have not played a musical instrument nor received music education. The summarized data can be found in Table 1. For the linear mixed-effects models, we treated music education as a categorical variable and musical experience as a numerical variable with years of musical practice of our participants as instrumentalists and/or vocalists (whichever was higher).

Euclidean Distance From GB Model Before and After Training

Figure 5 shows the mean distance from the GB model for all vowels before and after training for all 50 participants. A parametric t test for dependent means confirmed a significant difference between mean distance from the GB model before training (M = 0.23, SD = 0.13) and after training (M = 0.20, SD = 0.12), $t(499) = -5.46$, $p < .001$.

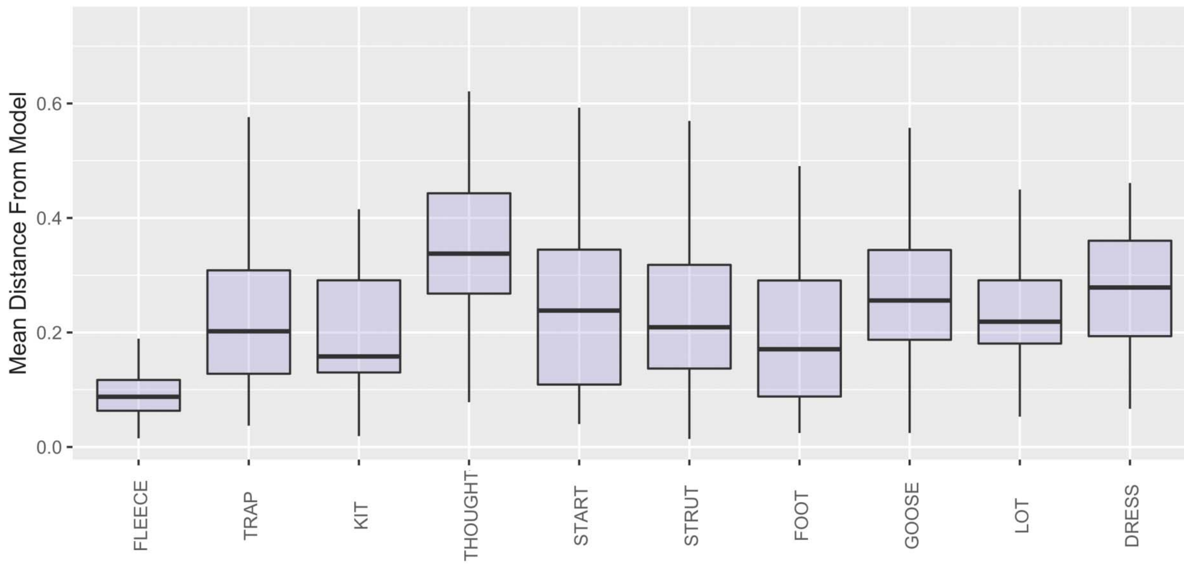
Table 1. Number of participants with self-reported musical experience.

Years	Music education	Musical instrument	Singing
12	2	2	10
9	—	3	—
6	2	6	2
3	—	1	1

Figure 6 shows the mean distance from the GB model for each individual vowel before training. A one-way analysis of variance reported a significant difference between the average measures of the 10 vowels, $F(9, 490) = 14.45$, $p < .01$, suggesting that the participants had already found some vowels to be more problematic in production than others. FLEECE produced by the participants was the closest vowel to the GB model (M = 0.09, SD = 0.05), while the most distant was THOUGHT (M = 0.34, SD = 0.13).

Figure 7 shows the mean distance from the GB model for each individual vowel after training. A one-way analysis of variance reported a similarly significant difference between the average measures of the 10 vowels, $F(9, 490) = 15.16$, $p < .01$, indicating that not all L2 vowels were equally mastered by the participants after training. Again, FLEECE produced by the participants was the closest vowel to the GB model and had little variation (M = 0.08, SD = 0.04). Next, the vowels with similar variation and mean distance from the model were TRAP, KIT, THOUGHT, START, and STRUT. More variation could be observed for back vowels (i.e., FOOT, LOT, and GOOSE, the latter exhibiting more significant variation [M = 0.28, SD = 0.15]). It is worth noting at this point that both FOOT and GOOSE vowels are often confused by Polish learners of English, as there is only one /u/ sound in their L1 vowel system. Furthermore, while Polish has a vowel similar to LOT in its phonetic inventory (i.e., the open-mid back /ɔ/, many Polish learners of English often substitute it with the General American /ɑː/), mostly due to their exposure to

Figure 6. Distance from General British model before training for individual 10 monophthongs.



American English at earlier stages of education and the predominant American English accents present in popular culture. Interestingly, the vowel, which, on average was the most extremely distant from the model, was DRESS ($M = 0.27$, $SD = 0.12$), although it should be a fairly easy vowel to acquire by Polish learners of English because, in their native vowel inventory, there exists a nearly equivalent open-mid front vowel /ɛ/. It would seem then that developing a new L2 vowel category would only entail raising the Polish /ɛ/ vowel slightly.

Figure 8 shows two contrasting participants producing GB vowels before training. P45 (left) managed to produce vowel sounds that were similar to the GB model, while the tokens of P29 (right) were further from the model. It is apparent that, while P45 has a more consistent and categorical distribution of vowel sounds, the same tokens for P29 often overlap or show a considerable degree of variation. Figure 9 shows the same two participants after their 1-year accent training. It can be observed that the former acquired a more categorical distribution of vowels with a relatively low degree of variation, while the latter still has apparent overlaps in their vowel system and a higher degree of variation for some vowels.

Figure 7. Distance from General British model after training for individual 10 monophthongs.

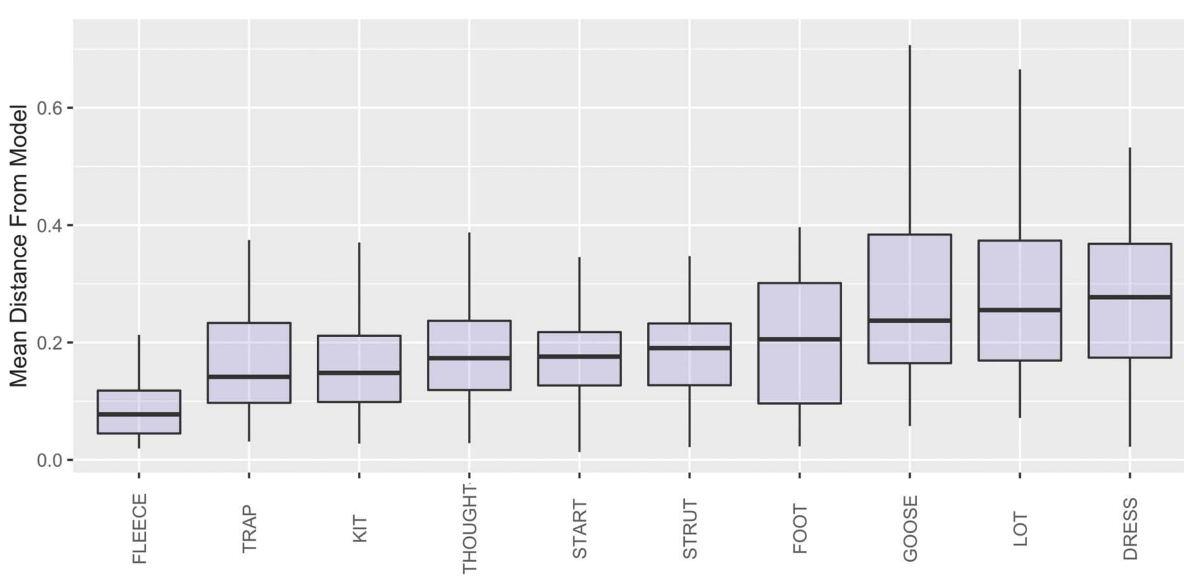
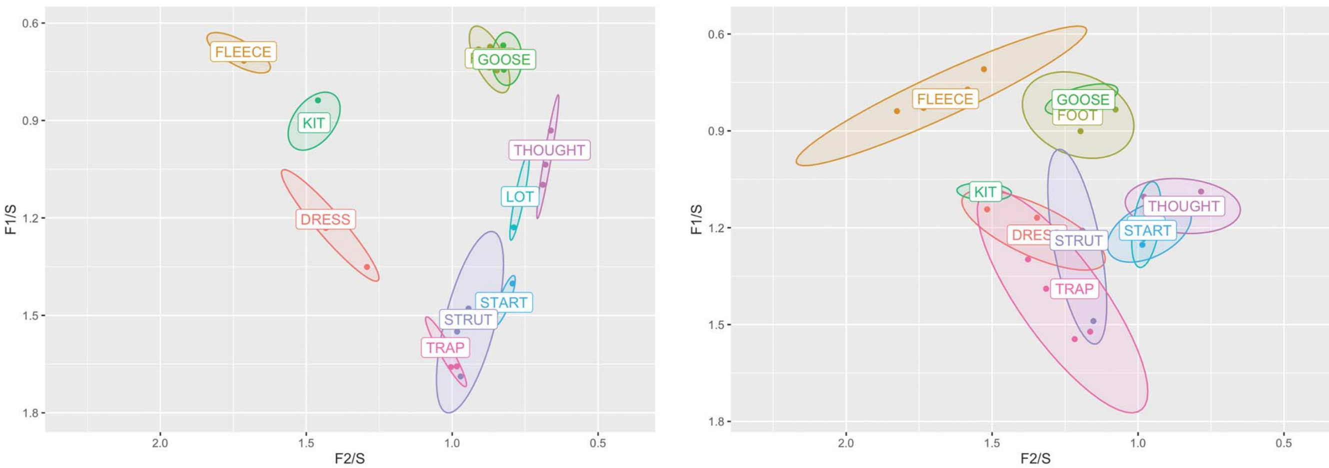


Figure 8. Example vowel formants before training for P45 (left) and P29 (right).



Closeness to Model Before Training

Prior to investigating participants' vowel production after training, we looked into the effects of individual variables on their results before the accent training course. A linear mixed-effects regression model was built to explain the closeness of vowel formants produced by the participants' to the model values before training, with speaker and vowel as random effects and gender, LexTALE score, pitch perception test result, melodic memory test result, rhythmic memory test result, music education, and years of musical experience as fixed effects. The results for 500 observations of 50 speakers and 10 vowels presented in Table 2 show a significant result for rhythmic memory (p = .004). A negative estimate (−.002) for that parameter indicates that a more accurate rhythmic memory is associated with a smaller distance from the pronunciation model. Therefore, participants who scored higher on that particular musical hearing test (i.e., were better at differentiating two rhythmic patterns) were also closer to the model. For example, participant P016 with the highest score (92%) in the rhythmic memory test was also the closest to the pronunciation model, with .164 in their mean closeness to the model before training. Conversely, participant P029 with one of the lowest results in this test (52%) was the furthest from the model before training (.374). We found no significant results for the pitch perception test results or the melodic memory test results. Likewise, we also found no relation between participants' closeness to model before training and their music education or musical experience. We reported no multicollinearity between the independent variables in the VIF scores (< 1.5).

For the estimated random effects, we found a significant difference between variance for speaker (.0001, SD = 0.01) and vowel groups (.0039, SD = 0.06), which suggests that the differences between vowels can be more informative

Figure 9. Example vowel formants after training for P45 (left) and P29 (right).

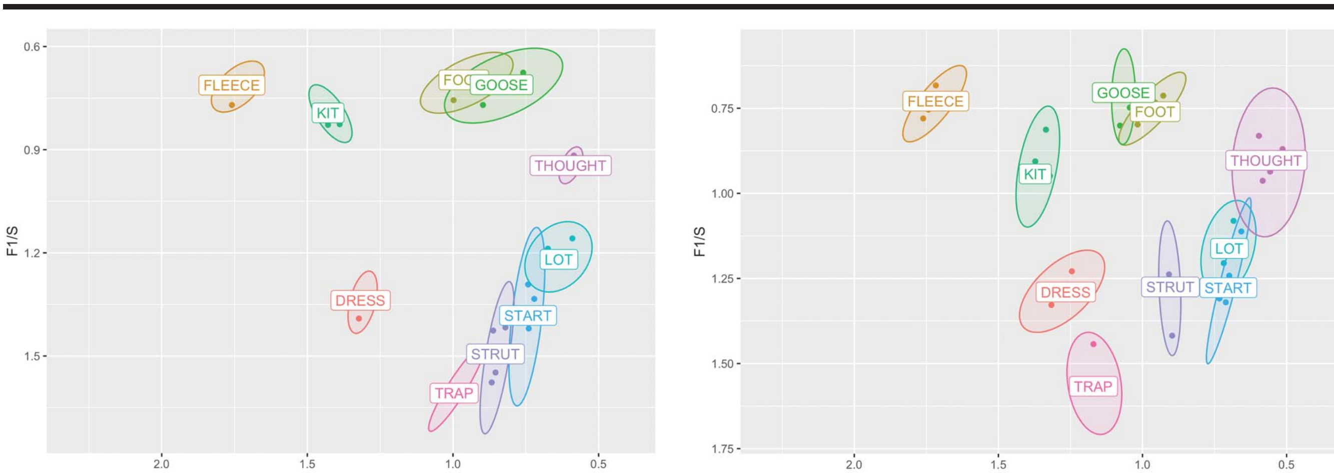


Table 2. Linear mixed-effects model results for closeness to model before training (500 observations, 50 speakers, 10 vowels).

Parameter	Estimate	*SE*	Test (df)	*p*
Intercept	0.290	0.07	4.14 (47.75)	< .001
Gender (M)	−0.006	0.02	−0.34 (41.99)	.733
LexTALE	0.000	0.00	0.30 (41.99)	.762
Pitch perception	0.000	0.00	1.28 (41.99)	.207
Melodic memory	0.000	0.00	1.10 (41.99)	.275
Rhythmic memory	**−0.002**	**0.00**	**−3.01 (41.99)**	**.004**
Music education	0.001	0.02	0.07 (41.99)	.945
Musical experience	0.000	0.00	0.38 (41.99)	.702

Note. *SE* = standard error; M = male; LexTALE = Lexical Test for Advanced Learners of English. Boldface indicates statistical significance ($p \leq .05$).

Table 3. Linear mixed-effects model results for closeness to model after training (500 observations, 50 speakers, 10 vowels).

Parameter	Estimate	*SE*	Test (df)	*p*
Intercept	0.186	0.07	2.82 (48.77)	.006
Gender (M)	0.018	0.02	1.12 (42.01)	.267
LexTALE	−0.000	0.00	−0.67 (42.00)	.506
Pitch perception	−0.000	0.00	−0.47 (42.20)	.641
Melodic memory	−0.000	0.00	−0.49 (42.15)	.628
Rhythmic memory	−0.000	0.00	−0.24 (43.16)	.810
Music education	0.022	0.02	1.02 (41.99)	.312
Musical experience	**−0.002**	**0.00**	**−2.12 (42.01)**	**.039**
Before training	**0.368**	**0.04**	**9.68 (489.1)**	**< .001**

Note. *SE* = standard error; M = male; LexTALE = Lexical Test for Advanced Learners of English. Boldface indicates statistical significance ($p \leq .05$).

than between speakers. We calculated separate linear regressions for each vowel to investigate the relation between rhythmic memory and participants' closeness to model before training. We reported a significant result with a negative estimate −.004 for TRAP, $F(1, 48) = 6.06$, $p = .017$, $R^2 = .09$, and a weak significant result with a negative estimate −.003 for FOOT, $F(1, 48) = 3.63$, $p = .062$, $R^2 = .05$. This is an interesting finding, since Polish students of English often struggle with the pronunciation of TRAP and substitute it with Polish /ɛ/ or /a/ (Weckwerth, 2011), while FOOT is often confused with GOOSE and perceived as Polish /u/ (Balas, 2018). Both vowels were also similarly challenging for our participants before training.

Closeness to Model After Training

To predict the closeness of vowel formants produced by the participants' to the model values after training, we built a linear mixed-effects regression model with speaker and vowel as random effects and gender, LexTALE score, pitch perception test result, melodic memory test result, rhythmic memory test result, music education, years of musical experience, and the closeness to the model before training as fixed effects. The results for 500 observations of 50 speakers and 10 vowels presented in Table 3 show a significant result with a negative estimate for musical experience (−.002, $p = .039$) and a very significant result with a positive estimate for the closeness to model before training (.368, $p < .001$). We reported no multicollinearity between the independent variables in the VIF scores (< 1.5). These results indicate that participants with more years of musical experience also achieved a more native-like pronunciation by producing similar vowels to their pronunciation teachers. For instance, participant P055 with the closest mean values after training (.125) also had 12 years of singing experience, while participant P022 who was the furthest from the model (.307) had no musical experience. We found no significant results for the musical hearing tests, possibly suggesting that either musical hearing is not an asset during explicit pronunciation training in a formal learning environment, or that the training itself is effectively flattening any potential influence of musical hearing on participants' progress. The results also suggest that years of practicing music, either by playing a musical instrument or singing, can be more informative than having formal music education, as we found no significant result for that parameter. However, it is important to point out that we only had four participants who reported attending music school. To further investigate this dependence, it would be required to have a considerable number of participants divided into two equal groups of formally trained musicians and amateur musicians, as well as an insight into their actual musical performance and motivation in their music education or practice.

Similarly to the previous model, we found a significant difference between variance for speaker (.0003, $SD = 0.02$) and vowel (.0024, $SD = 0.05$). We calculated separate linear regressions for each vowel to investigate the relation between musical experience and participants' closeness to model after training. We reported a significant result with a negative estimate −.008 for TRAP, $F(1, 48) = 6.23$, $p = .016$, $R^2 = .09$, and a weak significant result with a negative estimate −.006 for STRUT, $F(1, 48) = 4.37$, $p = .041$, $R^2 = .06$. As previously established, TRAP is a difficult vowel for Polish learners of English, and it is possible that musical practice can affect participants' progress in the acquisition of this particular vowel, while STRUT is also often substituted by Polish /a/ or mispronounced due to spelling pronunciation (Weckwerth, 2011).

Being able to produce vowels similar to the GB model prior to the actual training was a determining factor, confirming that participants who were relatively accurate in their pronunciation from the start were also closer to the model after training. A participant's mean distance from the GB model is expected to increase by .368 for every unit of distance from that model before training ($p < .001$) after controlling for the other variables. While there were participants who produced vowels similar to their pronunciation teachers both before and after training (e.g., P028 before = .192, after = .191), there were also participants who made a significant progress during training (e.g., P029 before = .374, after = .197). However, to determine the effect of musical hearing on the actual progress in the acquisition of L2 pronunciation, it would be required to have a study group comprising participants with fairly similar vowel qualities at the beginning of the study. This is why we decided not to measure the difference between the closeness to model before and after training as such, but rather focus on the participants' performance after training, including their closeness to model at the beginning of the accent training course as one of the explanatory variables.

We also calculated separate linear regressions for each vowel to investigate the relation between participants' closeness to model before and after training. We reported strong significant results with a positive estimate .856 for DRESS, $F(1, 48) = 59.93$, $p < .001$, $R^2 = .55$; .778 for GOOSE, $F(1, 48) = 34.1$, $p < .001$, $R^2 = .40$; .666 for FOOT, $F(1, 48) = 35.06$, $p < .001$, $R^2 = .41$; .311 for FLEECE, $F(1, 48) = 6.85$, $p = .012$, $R^2 = .11$; .262 for THOUGHT, $F(1, 48) = 10.16$, $p = .002$, $R^2 = .16$; and .241 for START, $F(1, 48) = 11.63$, $p = .001$, $R^2 = .18$. Interestingly, most of these vowels are considered as long (i.e., having a dimension not present in the Polish vowel system), while at the same time we found no significant results for KIT, TRAP, STRUT, and LOT (i.e., short vowels, which are often substituted with Polish vowel equivalents by Polish learners of English). This would suggest that learners who produced more native-like long vowels after training had been already familiar with these sounds, while other participants were still far from the model after the pronunciation course.

Finally, the results also show that the LexTALE result had no effect on the expected distance from the GB model, suggesting that a general language proficiency, especially relating to vocabulary, is not related to the acquisition of a native-like accent of English. However, a Pearson *r* test confirmed a positive, though weak correlation between the LexTALE results and the melodic memory ($r = .06$) and rhythmic memory ($r = .11$) tests, suggesting that a follow-up study to investigate the relationship between general language proficiency and specific aspects of musical hearing can be valuable.

Discussion

The study examined the acquisition of L2 vowels by 50 Polish advanced learners of English during a two-semester accent training course and the effect of musical hearing and musical experience on the estimated closeness to the GB pronunciation model. The experiment included an acoustic analysis of vowel formants before and after training, a series of musical hearing tests, and a questionnaire regarding musical experience. According to the results, L2 pronunciation is trainable in adult Polish learners of English, and specific aspects of musical hearing and musical experience can predict successful acquisition of L2 vowels before and after training. Although the learners' former proficiency in L2 pronunciation is the most robust predictor in acquiring a native-like accent, preexisting rhythmic memory is also positively associated with pretraining L2 pronunciation, while years of musical practice can predict better posttraining production of L2 vowels, even when pretraining pronunciation is held constant.

We predicted that the participants should produce L2 vowels closer to the pronunciation model after training. It turned out that there was a significant difference between the participants' vowels before and after training, suggesting that a 1-year pronunciation course can have an effect on L2 vowel acquisition and that L2 pronunciation is teachable in a formal academic context. Moreover, we observed that not all vowels were similarly easy to acquire by our participants: The formant values of front vowels FLEECE, KIT, and TRAP were the closest to the model after training, while the most distanced were back vowels GOOSE, FOOT, and LOT. Interestingly, the DRESS vowel was the most difficult to be acquired by our participants. The reason for this can be twofold. First, the DRESS vowel in English is a near-counterpart of the Polish vowel /ɛ/. Perceptually, they can be quite similar, especially considering the recent trend for the GB vowel DRESS to be produced as lower than in the more conservative normative accent labeled received pronunciation (i.e., British English pronunciation based on educated speech in southern England; cf. Cruttenden, 2014). Therefore, because perceptually both vowels /ɛ/ are so similar, the Polish students did not develop a new category for their English vowel. This is perhaps why the teachers usually do not devote a lot of time for discussing and practicing this vowel. Secondly, what could have added to the effect is that the traditional way of teaching the DRESS vowel in Polish higher education context by some teachers has been to instruct the students to produce the vowel higher than in Polish (i.e., in a more conservative way). When we investigated the formant values in the production of the DRESS vowel by the teachers, it turned out that indeed their DRESS vowel was produced higher than their Polish /ɛ/ vowel, similarly to the received pronunciation values found in Cruttenden (2014). This rather conservative feature of their accents could have been ignored by the students or consciously unlearnt, possibly due to their exposure to other contemporary English accents in the media (i.e., students did not incorporate a conservative language feature to their repertoire, if they had heard it produced differently by the native speakers). There also remains the question to what extent our results are extendable to other language pairs; while Polish learners often have problems with learning the TRAP vowel (Weckwerth, 2011) or the contrasts between FLEECE and KIT vowels (Rojczyk & Porzuczek, 2012), speakers with different L1s may find other contrasts more problematic.

We assumed that participants who received better results in the musical hearing tests should also produce vowels closer to their pronunciation model after training. While the test results assessing pitch perception and melodic memory were not associated with the closeness to the model, we reported that the rhythmic memory test scores were related to the shorter distance between the participants' formant values of vowels and the pronunciation model before training. Although language rhythm is commonly associated with suprasegmental features, studies have also suggested links between rhythm and segmental phonology, particularly relating to the temporal aspects of the tense–lax contrast in English (Schwartz, 2010). Indeed, rhythmic perception can have an important role in successful language acquisition (Jusczyk, 1999) and practicing musical rhythm can help achieve more native-like pronunciation and fluency (Llanes-Coromina et al., 2018). Former studies also confirmed the existence of shared neurocognitive resources for rhythm in music and speech (e.g., Magne et al., 2016), as well as successful use of musical rhythm in teaching L2 English prosody, particularly in practicing duration contrasts between stressed and unstressed syllables (Wang et al. 2016). Moreover, Milovanov et al. (2010) also noted in their study that choir members, who performed better in discriminating rhythmic patterns than nonmusicians, also produced fewer pronunciation errors than nonmusical university students. Since many Polish learners of English struggle with achieving native-like L2 pronunciation due to the complex English vowel system with its durational contrasts and vowel reduction, it is possible that rhythmic memory can help such learners achieve more native-like pronunciation through superior discrimination between tense and lax vowels. However, since our study was interested primarily in vowel quality, it would be valuable to see a follow-up study investigating the role of musical rhythm in the acquisition of L2 English vowels by Polish learners of English in terms of both vowel duration and vowel reduction.

We found no relation between participants' closeness to model before or after training and their pitch perception test results, although it is worth noticing that the reported high values for this test vary from similar previous studies (e.g., Amitay et al., 2006; Micheyl et al., 2006), even though all participants were identically instructed and used the same equipment. Possibly, participants reacted differently to the test due to the relatively long recording session, which preceded the musical hearing test and the weak results might have been caused by fatigue. Alternatively, the fact that the pitch perception test was conducted before the melodic memory test and rhythmic memory test might have affected the results. Finally, this is the first study that used this method on Polish learners of English, so perhaps the weak results stem either from insufficient music education in Polish schools (see Zwolińska, 2008) or from L1 influence on auditory processing, which was not yet tested across Polish learners of English (cf. Dawson et al., 2017). These results suggest that musical hearing is not one concept but that it comprises different abilities, which are also differently applicable to production skills.

We also assumed a potential relation between successful L2 vowel acquisition and musical experience. A significant result for this parameter suggests that participants who spent more years practicing music, either by playing a musical instrument or singing, also produced more native-like vowels after training. At the same time, we found no relationship between music education and successful L2 vowel acquisition before or after training. However, it is important to point out that only four participants reported attending music school, thus requiring further investigation in a more balanced sample. Since this study was conducted among Polish students of English in an academic context, we had no control over their former music education or musical experience. One important parameter that should be considered is the starting age of music education and training, as it can be related with successful language acquisition at an early stage (Brandt et al., 2012), as well as helping young learners with hearing impairment (Torppa & Huotilainen, 2019). This could be a potential point of departure for future studies.

The strongest observable estimate in our study was the closeness to model before training (i.e., participants who managed to have acquired more native-like pronunciation before the two-semester accent training course were also more successful afterwards). Consequently, the effects of the intensive formal instruction in the form of pronunciation training did not factor out all the differences between the participants in this study. Even though the participants were a fairly homogeneous group in terms of their age, LexTALE results measuring their overall language proficiency, as well as studying the same degree, English, which required from them similar secondary final exam results, they varied in terms of their pronunciation before the study. Moreover, all participants confirmed that they had not obtained any pronunciation instruction before the study. This means that, given the differences between the speakers before the training, the learners must have ways of assimilating elements of L2 pronunciation outside the context of formal instruction.

Finally, we observed different relations between the closeness to model for individual vowels and the significant variables from the models. The most interesting finding was for TRAP, a considerably difficult vowel for Polish learners of English who frequently substitute it with /ɛ/ or /a/ (Weckwerth, 2011). While we found no relation between the production of TRAP before and after training, we observed a significant result for rhythmic memory and musical experience, indicating that both musical hearing and musical practice can be strongly associated with the successful acquisition of this particular sound. Rhythmic memory was also found as a significant parameter explaining the closeness to model for FOOT, another difficult vowel, commonly confused with GOOSE, and exhibiting considerable variation among Polish learners of English. For musical experience, we also reported a significant result for STRUT, a problematic vowel that is often mispronounced due to spelling pronunciation. Lastly, we observed that the majority of vowels produced after training with a significant result for the "before training" parameter were long vowels (FLEECE, START, GOOSE, THOUGHT), suggesting that participants who had already acquired these sounds were closer to the model after training, while participants who struggled with the tense–lax contrast before the pronunciation course had still some problems with these vowels. These results suggest that musical hearing and musical experience are not only connected to a more native-like L2 vowel system as a whole, but can be related to specific vowels that pose considerable problems for Polish learners of English.

It is difficult at this point to assess the teaching implications of our results. On the one hand, we were investigating the talent or the aptitude aspect in the process of L2 phonology acquisition because it had been clearly shown in many studies before that motivation and formal instruction both play a significant role in this process. On the other hand, both musical and linguistic perception are trainable. There is a growing body of evidence (Baese-Berk, 2019) suggesting that production can shape perception, precisely in the context of foreign-language learning. It is better to expose learners of English to real-life accent models as early as possible in the classroom setting to help them in successful acquisition of L2 pronunciation (Darcy et al., 2012). Moreover, while this study was primarily focused on the effect of musical hearing on L2 pronunciation, it is also possible that an intensive accent training course could affect musical hearing, as many exercises during the course rely on exposure, including ear training exercises and auditory discrimination tests, which, in turn, can influence the learners' awareness to speech sounds and musical sounds, as well. Furthermore, in recent years, there has been an increasing interest in the importance of learner's motivation over innate talent. While the concept of an inborn gift for pronunciation has been widely researched (e.g., Dogil & Reiterer, 2009; Jilka et al., 2011), the relevance of hard work still remains understudied. Grit, as defined by Duckworth et al. (2007), is the "perseverance and passion for long-term goals," and it can be significant in predicting the learner's success. Therefore, it would be of great value to develop reliable tools for measuring not only learners' talents but also their motivation.

Admittedly, the concept behind this article is based on the comparison of two different domains, which are perception and production. Moreover, the conceptual and statistical model used is multimodal and may not account for all aspects of vowel acquisition in advanced learners of English. Yet speech is a complex, multimodal, and context-dependent means of communication. The proposed model hopefully contributes to a better understanding of the mechanisms involved in the acquisition of L2 vowels and the role of specific aspects of musical hearing and musical experience that play in this process.

Conclusions

This study reported an experiment investigating whether preexisting musical hearing skills and musical experience can predict more native-like production of L2 vowels by Polish advanced learners of English before and after formal accent training. The results show that rhythmic memory is positively associated with more native-like L2 pronunciation before training, while years of musical experience can predict more accurate L2 pronunciation after training. Overall, the mean Euclidean distance between the participants' vowels and the model vowels produced by their pronunciation teachers decreased after training, but the change was more distinct for participants who performed better in the rhythmic memory test or had some musical experience as instrumentalists or vocalists. At the same time, we found no significant relation between the mean distance from the model and music education, suggesting that attending music school on its own might not be a predicting factor in the acquisition of native-like pronunciation. Finally, we found a strong relationship between the results before and after training, indicating that participants' former proficiency was related with their improved performance after the accent training course. Thus, while musical hearing and musical experience can play a role in the acquisition of L2 English vowels, individuals' L2 pronunciation before accent training seems to be the determining factor in explaining the closeness to the model after accent training.

Acknowledgments

This research was supported by the National Science Centre in Poland, Grant 2014/15/N/HS2/03865. Principal investigator: Mateusz Jekiel. Recipient: Adam Mickiewicz University in Poznań, Poland. The authors would like to express their gratitude to Professor Piotr Gąsiorowski for his assistance in the early stages of the project, as well as Professor Robert Lew and Kacper Łodzikowski for their help in the data analysis.

References

Amitay, S., Irwin, A., & Moore, D. (2006). Discrimination learning induced by training with identical stimuli. *Nature Neuroscience, 9,* 1446–1448. https://doi.org/10.1038/nn1787

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83*(2), 111–130. https://doi.org/10.1016/S0022-0965(02)00124-8

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics, 81*(4), 981–1005. https://doi.org/10.3758/s13414-019-01725-4

Balas, A. (2018). English vowel perception by Polish advanced learners of English. *Canadian Journal of Linguistics, 63*(3), 309–338. https://doi.org/10.1017/cnj.2018.5

Barrett, K. C., Ashley, R., Strait, D. L., & Kraus, N. (2013). Art and science: How musical training shapes the brain. *Frontiers in Psychology, 4,* 713. https://doi.org/10.3389/fpsyg.2013.00713

Baudeat, A. (2017). The emergence of language in the hominin lineage: Perspectives from fossil endocasts. *Frontiers in Human Neuroscience, 11,* 427. https://doi.org/10.3389/fnhum.2017.00427

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition, 77*(1), 1–10. https://doi.org/10.1016/j.bandc.2011.07.00

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (Version 5.4.01) [Computer software]. http://www.praat.org/

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology, 3,* 327. https://doi.org/10.3389/fpsyg.2012.00327

Brown, S. (2001). The "musilanguag" model of music evolution. In S. Brown, B. Merker, & N. Wallin (Eds.), *The origins of music* (pp. 271–301). MIT Press.

Brown, S., Martinez, M., & Parsons, L. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 23*(10), 2791–2803. https://doi.org/10.1111/j.1460-9568.2006.04785

Casillas, J. (2015). Production and perception of the /i/-/I/ vowel contrast: The case of L2-dominant early learners of English. *Phonetica, 72*(2–3), 182–205. https://doi.org/10.1159/000431101

Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences, 3*(2), 923–940. https://doi.org/10.3390/brainsci3020923

Council of Europe. (2011). *Common European Framework of Reference for Languages: Learning, teaching, assessment*.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). Routledge. https://doi.org/10.4324/9780203784969

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2011* (pp. 93–108). Iowa State University.

Dawson, C., Aalto, D., Šimko, J., Vainio, M., & Tervaniemi, M. (2017). Musical sophistication and the effect of complexity on auditory discrimination in Finnish speakers. *Frontiers in Neuroscience, 11,* 213. https://doi.org/10.3389/fnins.2017.00213

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. John Benjamins. https://doi.org/10.1075/lllt.42

Dogil, G. & Reiterer, S. (Eds.). (2009). *Language talent and brain activity*. De Gruyter Mouton. https://doi.org/10.1515/9783110215496

Dolman, M., & Spring, R. (2014). To what extent does musical aptitude influence foreign language pronunciation skills? A multi-factorial analysis of Japanese learners of English. *World Journal of English Language, 4*(4), 1–11. https://doi.org/10.5430/wjel.v4n4p1

Draxler, J. Ch., & Jänsch, K. (2019). *SpeechRecorder — A universal platform independent multi-channel audio recording Software* [Computer program] (Version 4.4.50). Retrieved September 1, 2019, from https://www.bas.uni-muenchen.de/forschung/Bas/software/speechrecorder/

Duckworth, A., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Fabricius, A. H., Watt, D., & Johnson, D. E. (2009). A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change, 21*(3), 413–435. https://doi.org/10.1017/S0954394509990160

Fadiga, L., Craighero, L., & D'Ausillo, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences, 1169,* 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Fenk-Oczlon, G. (2017). What vowels can tell us about the evolution of music. *Frontiers in Psychology, 8,* 1581. https://doi.org/10.3389/fpsyg.2017.01581

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 229–273). York Press. https://doi.org/10.1177/002383099503800102

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. Hualde (Eds.), *Laboratory phonology* (Vol. 9, pp. 353–380). Mouton de Gruyter.

Fonseca-Mora, M., Toscano-Fuentes, C., & Wermke, K. (2011). Melodies that help: The relation between language aptitude and musical intelligence. *Anglistik International Journal of English Studies, 22*(1), 101–118.

Franklin, M., Moore, K., Yip, C., & Jonides, J. (2008). The effects of musical training on verbal memory. *Psychology of Music, 36*(3), 353–365. https://doi.org/10.1177/0305735607086044

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 23*(27), 9240–9245. https://doi.org/10.1523/JNEUROSCI.23-27-09240.2003

Galińska-Brawata, A., & Rybińska, P. (2017). The relationship between the production of word stress and musical abilities in Polish learners of English. *Research in Language, 15*(3), 265–283. https://doi.org/10.1515/rela-2017-0015

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience, 29*(10), 3019–3025. https://doi.org/10.1523/JNEUROSCI.5118-08.2009

Ikeda, S., Shibata, T., Nakano, N., Okada, R., Tsuyuguchi, N., Ikeda, K., & Kato, A. (2014). Neural decoding of single vowels during covert articulation using electrocorticography. *Frontiers in Human Neuroscience, 8,* 125. https://doi.org/10.3389/fnhum.2014.00125

Isbell, D. (2016). The perception-production link in L2 phonology. *MSU Working Papers in SLS 2016, 7,* 57–67.

Jilka, M., Lewandowski, N., & Rota, G. (2011). Investigating the concept of talent in phonetic performance. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (pp. 171–180). Peter Lang.

Jusczyk, P. (1999). Narrowing the distance to language: One step at a time. *Journal of Communication Disorders, 32*(4), 207–222. https://doi.org/10.1016/S0021-9924(99)00014-3

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5,* 1246. https://doi.org/10.3389/fpsyg.2014.01246

Kendall, T., & Thomas, E. R. (2010). *Vowels: Vowel manipulation, normalization, and plotting in R*. R package. http://cran.r-project.org/web/packages/vowels/index.html

Kolinsky, R., Pascale Lidji, P., Peretz, I., Besson, M., & Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition, 112*(1), 1–20. https://doi.org/10.1016/j.cognition.2009.02.014

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLOS ONE, 10*(11). https://doi.org/10.1371/journal.pone.0141069

Lappe, C., Herholz, S., Trainor, L., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *The Journal of Neuroscience, 28*(39), 9632–9639. https://doi.org/10.1523/JNEUROSCI.2254-08.2008

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758%2Fs13428-011-0146-0

Lidji, P., Jolicoeur, P., Régine Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology, 121*(4), 533–541. International Speech Communication Association. https://doi.org/10.1016/j.clinph.2009.12.018

Llanes-Coromina, J., Prieto, P., & Rohrer, P. (2018). Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *9th International Conference on Speech Prosody* (pp. 498–502). https://doi.org/10.21437/SpeechProsody.2018-101

Maddieson, I. (2013). Vowel quality inventories. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

Magne, C., Jordan, D., & Gordon, R. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language, 153–154,* 13–19. https://doi.org/10.1016/j.bandl.2016.01.001

Mandell, J. (2009). *Electronic music and medical education*. Accessed September 1, 2019, from http://jakemandell.com

Mandell, J., Schulze, K., & Schlaug, G. (2007). Congenital amusia: An auditory-motor feedback disorder. *Restorative Neurology and Neuroscience, 25*(3–4, 334), 323.

Masataka, N. (2007). Music, evolution and language. *Developmental Science, 10*(1), 35–39. https://doi.org/10.1111/j.1467-7687.2007.00561.x

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In F. Lacerda (Eds.), *Proceedings of the 18th Conference of the International Speech Communication Association* (pp. 498–502). International Speech Communication Association. http://doi.org/10.21437/Interspeech.2017-1386

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P. A. (2010). Foreign language pronunciation skills and musical aptitude: a study of Finnish adults with higher education. *Learning and Individual Differences, 20*(1), 56–60. http://dx.doi.org/10.1016/j.lindif.2009.11.003

Mithen, S. (2005). *The singing neanderthals: The origins of music, language, mind and body*. London Weidenfeld & Nicolson.

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

Nikolsky, A. (2015). Evolution of tonal organisation in music mirrors symbolic representation of perceptual reality. Part-I: Prehistoric. *Frontiers in Psychology, 6,* 1405. https://doi.org/10.3389/fpsyg.2015.01405

Ott, C. G., Langer, N., Oechslin, M. S., Meyer, M., & Jäncke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Frontiers in Psychology, 2,* 195. https://doi.org/10.3389/fpsyg.2011.00195

Pastuszek-Lipińska, B. (2008). Musicians outperform nonmusicians in speech imitation. *Lecture Notes in Computer Science, 4969,* 56–73. https://doi.org/10.1007/978-3-540-85035-9_4

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Patel, A. D. (2012). Language, music, and the brain: A resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 204–223). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199553426.003.0022

Picavet, F., Auberge, V., & Rossato, S. (2012). Can a guided rhythmic approach contribute to the oral performance of learners of L2 English? A case study. In M. Busa & A. Stella (Eds.), *Methodological perspective on second language prosody. Papers from ML2P 2012* (pp. 73–77). Cleup.

Reddy, S., & Stanford, J. (2015). A web application for automated dialect analysis. In M. Gerber, C. Havasi, & F. Lacatusu (Eds.), *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 72–75). Association for computational Linguistics. https://doi.org/10.3115/v1/N15-3015

Rojczyk, A., & Porzuczek, A. (2012). Selected aspects in the acquisition of English phonology by Polish learners—Segments and prosody. In D. Gabryś-Barker (Ed.), *Readings in second language acquisition* (pp. 93–120). Wydawnictwo Uniwersytetu Śląskiego.

Roncaglia-Denissen, M. P., Roor, D. A., Chen, A., & Sadakata, M. (2016). The enhanced musical rhythmic perception in second language learners. *Frontiers in Human Neuroscience, 10,* 288. https://doi.org/10.3389/fnhum.2016.00288

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2014). *FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2*. http://doi.org/10.5281/zenodo.22281

Russo, F. A., Vuvan, D. T., & Thompson, W. F. (2019). Vowel content influences relative pitch perception in vocal melodies. *Music Perception: An Interdisciplinary Journal, 37*(1), 57–65. http://doi.org10.1525/mp.2019.37.1.57

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*(8), 511–514. https://doi.org/10.1111/j.0956-7976.2004.00711.x

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5*(7), 688–694. https://doi.org/10.1038/nn871

Schwartz, G. (2010). Rhythm and vowel quality in accents of English. *Research in Language, 8,* 135–147. https://doi.org/10.2478/v10015-010-0011-8

Seashore, C. E., Lewis, D., & Saetveit, J. (1960). *Seashore measures of musical talents*. The Psychological Corporation.

Slevc, L., & Miyake, A. (2006). Individual differences in second language proficiency: Does musical ability matter. *Psychological Science, 17,* 675–681. https://doi.org/10.1111%2Fj.1467-9280.2006.01765.x

Sobkowiak, W. (2008). *English phonetics for Poles* (3rd ed.). Wydawnictwo Poznańskie.

Strait, D., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language, 123*(3), 191–201. https://doi.org/10.1016/j.bandl.2012.09.001

Swaminathan, S., & Schellenberg, E. (2019). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000798

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing musical experience and language development. *Trends in Neuroscience, 29*(7), 382–390. https://doi.org/10.1016/j.tins.2006.06.003

Tankus, A., Fried, I., & Shoham, S. (2012). Structured neuronal encoding and decoding of human speech features. *Nature Communications, 3,* 1015. https://doi.org/10.1038/ncomms1995

Torppa, R., & Huotilainen, M. (2019). Why and how music can be used to rehabilitate and develop speech and language skills in hearing-impaired children. *Hearing Research, 380,* 108–122. https://doi.org/10.1016/j.heares.2019.06.003

Trehub, S. E., Schellenberg, G., & Kamenetzky, S. B. (1999). Infants' and adults' perception of scale structure. *The Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 965–975. https://doi.org/10.1037//0096-1523.25.4.965

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

Wang, H., Mok, P., & Meng, H. (2016). Capitalizing on musical rhythm for prosodic training in computer-aided language learning.

Computer Speech and Language, 37, 67–81. https://doi.org/10.1016/j.csl.2015.10.002
Weckwerth, J. (2011). English TRAP vowel in advanced Polish learners: Variation and system typology. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2110–2113). City University of Hong Kong.
Zhang, A., Feng, H., Wang, S., & Dang, J. (2016). *Relationship between perception and production of English vowels by Chinese English learners*. 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), Tianjin (pp. 1–5). https://doi.org/10.1109/ISCSLP.2016.7918479
Zwolińska, E. A. (2008). Polish fundamentals of music curriculum. *The GIML Audea, 13*(1), 7–8.
Zybert, J., & Stępień, S. (2009). Musical intelligence and foreign language learning. *Research in Language, 7*. https://doi.org/10.2478/v10015-009-0007-4