Poznań Studies in Contemporary Linguistics 58(2), 2022, pp. 315–340

doi: 10.1515/psicl-2022-0016

L2 RHYTHM PRODUCTION AND MUSICAL RHYTHM PERCEPTION IN ADVANCED LEARNERS OF ENGLISH

MATEUSZ JEKIEL
Adam Mickiewicz University, Poznań
mjekiel@amu.edu.pl

ABSTRACT

The aim of this research is to investigate the relation between musical aptitude and the acquisition of L2 rhythm by Polish advanced learners of English. A longitudinal study was conducted among 50 Polish students of English reading the "Please Call Stella" passage before and after an intensive two-semester accent training course supplemented by an extensive practical course in English phonetics and phonology. Participants also completed two musical hearing tests (Mandell 2009) and a survey on musical experience. Automated alignment was performed in DARLA (Reddy and Stanford 2015) and reviewed in Praat (Boersma and Weenink 2019). We compared the rhythm metrics calculated in Correlatore (Mairano and Romano 2010) before and after training and juxtaposed them against the pronunciation teachers' results. We reported a significant difference between the scores for vocalic intervals across all rhythm metrics, indicating that participants' produced higher vocalic variation after training, more similar to their teachers. However, we observed no significant relationship between the participants' rhythm metric scores and their musical hearing test scores or musical experience, suggesting that musical aptitude might not play a crucial role in the L2 rhythm production in a formal academic learning environment.

KEYWORDS: L2 rhythm; rhythm metrics; accent training; language and music; musical aptitude.

1. Introduction

Linguists and musicologists agree that the relationship between language and music can be observed on many levels: (1) both are specific to humans and appear in every society throughout history (Nettl 2000); (2) they share the same neural resources responsible for processing linguistic and musical structures (Koelsch et al. 2002); (3) both have tone, melody, and rhythm, as well as measurable duration, frequency, and timbre (Fadiga et al. 2009); (4) and they are both constructed out of syntactic sequences of sounds that can be used to produce an infinite number of possible hierarchically structured signals (Fenk-Oczlon 2009). These similarities led to a number of studies interested in establishing whether musical skills can be related to language skills. Both musical aptitude, i.e. the "potential to achieve in music" (Gordon 1989), as well as musical experience, i.e. music education and performance, have been connected to language skills (Magne et al. 2006). Musical aptitude can positively influence processing of speech, both in terms of perception and production (Chobert and Besson 2013); as well as phonological awareness (Culp 2017) and successful second language learning (Milovanov 2004). Formal music education and musical training can also have a positive effect on the brain mechanisms related to language skills (Wong et al. 2007), although it has also been argued that the majority of such studies do not derive their data from experimental studies (see Patel 2008 for fuller treatment).

Recent research suggests that musical aptitude can also have a positive effect on the acquisition of L2 pronunciation, including both segmental and suprasegmental features. In a study by Milovanov et al. (2008), children with superior musical skills acquired advanced English pronunciation, while Slevc and Miyake (2006) confirmed a link between musical aptitude and second language listening discrimination and speech production among Japanese 11-year-old learners of English. In the case of adult language learners, Milovanov et al. (2010) found a relation between musical aptitude and pronunciation skills in Finnish learners of English. Musical aptitude can also affect direct speech imitation (Pastuszek-Lipińska 2008), which constitutes a learning task, as well as accent faking (Coumel et al. 2019), which relies on retrieving previously stored phonetic representations. The above-mentioned studies suggest that investigating the exact interactions between specific aspects of musical aptitude, i.e. pitch perception, melodic memory, and rhythm perception, and specific aspects of L2 pronunciation, i.e. segmental and suprasegmental phonetics, should be regarded as essential in the process of understanding the precise relationship between these two domains.

While language and music share some commonalities with regard to rhythm, recent research still remains inconclusive. Both speech and music have complex hierarchical rhythmic structures and similarly use pitch and duration to mark group boundaries, albeit in a different manner (Patel 2008). Several studies confirmed the existence of shared neurocognitive resources for rhythm in speech and music (e.g. Magne et al. 2016), suggesting that rhythmic grouping in music can be considered as an offshoot of prosodic grouping in speech. Some researchers even suggested that musical rhythm can be a reflection of the musicians' language rhythm, based on similar differences found between the rhythm metric scores for the composer's native language and their instrumental music, e.g. in the case of English and French speech and early 20th century English and French classical music (Patel and Daniele 2003), although these results were later regarded as questionable due to the complex nature of classical music rhythmic patterns and the effect of sample selection on the rhythm metric scores (London and Jones 2010).

This study is particularly interested in the relationship between L2 rhythm and musical rhythm. According to Jusczyk (1999), rhythmic perception plays an important role in successful language acquisition and is related to phonological awareness (Moritz et al. 2013). Llanes-Coromina et al. (2018) also argue that practicing musical rhythm can improve L2 pronunciation, specifically in terms of perceived comprehensibility and fluency, which are often related to perceived language rhythm. Roncaglia-Denissen et al. (2016) suggest an alternative relationship, claiming that second language learners can develop enhanced musical rhythmic perception, especially if the learners' L1 rhythmic properties significantly differ from their L2. In the study, adult Mandarin, Turkish, and Dutch learners of English were compared with Turkish monolinguals in terms of their musical skills using a melodic and rhythmic aptitude test (Wallentin et al. 2010), similar in assumptions and procedures to the ones used in this study. The authors reported a significant difference in the rhythmic aptitude test scores between monolinguals and bilinguals, as well as between the learners whose L1 rhythm is traditionally considered as different than English, e.g. Mandarin learners of English scored better than their Dutch counterparts, although the latter still scored better than Turkish monolinguals. However, to the best of our knowledge, there is still no study that investigates the relationship between musical aptitude and the acquisition of L2 rhythm using language rhythm metrics. This study will try to enrich the current knowledge on the relation between L2 rhythm and musical rhythm, using quantitative rhythm metrics and musical aptitude measurements.

1.1. Language rhythm and rhythm metrics

Although there is a plethora of studies covering various intricacies of language rhythm, there is still no universally accepted definition of this phenomenon. According to Dellwo (2006), rhythm is the systematic organisation of prominent and less prominent speech units in time, where speech units are defined as syllables or vocalic intervals, while prominence as longer duration, stronger intensity or higher frequency. According to Gut (2012), speech rhythm can be regarded as the temporal organisation of language, a periodic recurrence of events or a structural property of a language, which allows to categorise it to a particular rhythm class or locate it on a rhythmic continuum. As all languages are rhythmically organised, researchers have tried for a considerable period of time to label every language in respect to their rhythmic features with various outcomes.

The first attempt to classify languages according to their rhythm was made by Pike (1945) using the concept of isochrony, i.e. the rhythmic division of time into equal portions, which was later supported by a physiological explanation found in Abercrombie (1967). According to this hypothesis, languages can be categorised either as syllable-timed, i.e. having syllables of equal duration (usually Romance languages, e.g. French, Italian, Spanish); or stress-timed, i.e. having patterns of equal duration between stressed syllables (usually Germanic languages, e.g. English, German, Dutch). While such a dichotomy can be regarded as structurally elegant, later studies have shown that many languages are not easily subject to such classification (Ramus et al. 1999), although some perceptual studies also confirmed that listeners can indeed discriminate between selected stress-timed and syllable-timed languages (Nazzi et al. 1998, Ramus et al. 2003). Therefore, there was a need to find acoustic correlates of language rhythm in the speech signal through experimental studies. Roach (1982) suggested that languages categorised as stress-timed would need to have significant variation between stressed and unstressed syllables and rely on vowel reduction to keep the stressed syllables relatively salient, while syllable-timed languages would need to have fairly equal syllable length.

Following these assumptions, Dauer (1983) proposed that rhythm is a gradient feature of a language, rather than an absolute one, and results from a combination of phonetic, phonological, lexical and syntactic facts, such as syllable structure, word stress and vowel reduction. The model proposed by Dauer is also considered as the precursor for the development of rhythm metrics, which allow to conduct quantitative research into language rhythm and measure the rhythmic differences between languages belonging to the conventional rhythmic categories. These rhythm metrics rely on the timing relations between vocalic and consonantal[1] intervals, assuming that language rhythm is

[1] Some studies use the term "intervocalic" instead of "consonantal". For consistency with rhythm metric names, we use the latter.

primarily determined by vocalic and consonantal variation. For example, stress-timed languages such as English or German allow vowel reduction and complex consonant clusters, which lead to higher variation of content of vocalic and consonantal intervals, while syllable-timed languages usually do not have these features, and consequently, have lower vocalic and consonantal variation.

The first rhythm metrics proposed by Ramus et al. (1999) were used to divide speech into vocalic and consonantal intervals in order to calculate the proportion of these intervals within a sentence and their standard deviation. %V and %C represent the timing proportions of vocalic and consonantal intervals in a single utterance, while ΔV and ΔC (also referred to as Vdev and Cdev respectively) stand for the standard deviations of these intervals. The results of these metrics for several languages corresponded to the traditional rhythm categories proposed by Pike (1945) and Abercrombie (1967). In a subsequent study by Ramus et al. (2003), these metrics have been validated in a perceptual study, where participants listening to modified speech with degraded segments and intonation could discriminate between English and Polish, which had different %V and ΔC scores, but couldn't differentiate between English and Dutch, which had similar rhythm metric scores.

Although %V is regarded as a fairly reliable predictor of rhythm types, ΔV and ΔC results can be easily affected by speech rate variation. To resolve this issue, Dellwo and Wagner (2003) and Dellwo (2006) introduced the rate-normalised metrics VarcoV and VarcoC in order to normalise the results across fluctuating speech rates by means of calculating the standard deviation of vocalic/consonantal interval duration, dividing it by the mean vocalic/consonantal interval duration and multiplied by 100. Dellwo and Wagner (2003) suggested that VarcoC can be more accurate than ΔC in discrimination between stress-timed and syllable-timed languages, while White and Mattys (2007) confirmed that VarcoV can be more precise than ΔV in capturing rhythmic differences between languages. While normalisation can provide more accurate results by neutralising differences in speech rate, Mairano and Romano (2011) also point out that this can potentially neglect relevant phenomena connected to speech tempo, thus it is recommended to use both raw and normalised rhythm metrics in order to obtain a more reliable and complete profile of the investigated material.

Another series of rhythm measurements was proposed by Low et al. (2000) and Grabe and Low (2002). The PVI (Pairwise Variability Indices) are measures of temporal patterning in speech used for comparing languages by means of calculating the degree of durational contrast between successive elements in a sequence. These metrics can be used to compare adjacent vocalic (PVI-V) or consonantal (PVI-C) intervals, providing the mean of the differences between these successive intervals divided by their sum. The PVI measures were designed respectively to reflect alternating stressed syllables and consonant cluster variation, which are considered as the two most important features of stress-timed languages. While a higher vocalic PVI result indicates higher variation in vowel duration and the presence of vowel reduction, a higher consonantal PVI result implies higher variation of consonantal clusters. The nPVI (normalised Pairwise Variability Indices) measures durational contrasts between consecutive elements in a sequence and normalises them for fluctuations in speech rate. Specifically, this rhythm metric calculates the variation in length of neighbouring pairs of vocalic and consonantal intervals and averages these results over the entire speech. Similarly to the rate-normalised VarcoV and VarcoC, nPVI-V and nPVI-C can avoid the distortion of results due to speech rate variation. These metrics were successfully used to study the rhythmic differences between accents of English, confirming that Singaporean English is more syllable-based than stress-based British English (Low et al. 2000). Although the difference between the two accents is relatively small when compared with other languages, it shows that rhythm metrics can be used to detect finer rhythmic characteristics both across languages and accents. Another method for normalising the PVI scores is the CCI (Control/Compensation Index) in Bertinetto and Bertini (2008), which divides the duration of each interval by the number of segments in it. Similarly to the PVI measures, CCI-V is used for vocalic intervals and CCI-C for consonantal intervals respectively.

In order to guarantee reliable and reproducible rhythm metric results, certain measures need to be implemented. First, it is important to rely on a consistent segmentation procedure, as the segmentation choices can affect the results (Mairano and Romano 2011). Automatic segmentation should provide more uniform results and facilitate the preparation of larger speech samples, as previous studies relied on a limited number of speakers due to time-consuming manual segmentation. Second, the use of one type of speech data is crucial, as rhythm metric results can significantly vary between spontaneous and read speech (Dellwo and Wagner 2003, Arvaniti 2012). Moreover, the structure of the reading text alone can influence the results, as Gibbon (2003) confirmed that the same ratio of alternating patterns and monotonic geometric series can yield the same rhythm metric scores. Finally, a common issue in studies on speech rhythm stems from speech rate, as the speed of articulation can affect the results of some rhythm metrics and reveal that rhythmic variability within a particular language can be as significant as between languages (Arvaniti 2012). For example, this can be a result of an unnatural, slowed down manner of reading a text by L2 speakers. Therefore, it is crucial to either control for speech rate or avoid comparing L2 speech to L1, as it might yield distorted results.

1.2. Rhythm in English and Polish

English has been traditionally classified as a prototypical stress-timed language, primarily due to its complex syllable structure, vowel reduction and variable stress patterns. On the other hand, the status of Polish rhythm category has been changing over the years. Rubach and Booij (1985) argued that Polish is an atypical case of a stress-timed language, while Avery and Ehrlich (1992) claimed that it should be regarded as a syllable-timed language due to its lack of vowel reduction. Alternatively, Polish has been categorised as a mixed-type language, occupying an intermediate position and exhibiting features of both stress-timed languages (variable and complex syllable types) and syllable-timed languages (lack of phonological vowel reduction) (Nespor 1990, Grabe and Low 2002). In a recent study by Mairano and Romano (2011), Polish was also considered as a rather isolated case across other European languages, having both low vocalic variability due to the lack of vowel reduction and high consonantal variability caused by complex consonant clusters. Due to these measurable differences between Polish and English rhythm, it should be possible to observe non-native rhythm metric scores in Polish learners of English when they are speaking English. Specifically, we expect that the L2 English rhythm metric scores will be closer to L1 Polish than L1 English due to inconsistent vowel reduction and vowel duration, the two key features that significantly affect rhythm metric scores and are considerably different in Polish and English.

1.3. Teaching second language rhythm

While language rhythm is considered to be one of the first aspects of language acquired by infants (Nazzi et al. 1998), the acquisition of L2 rhythm can be problematic, especially for adult learners, and contributes to the perception of foreign accent (Barry 2007). A number of studies have confirmed the influence of prosodic features on learner’s intelligibility and comprehensibility (Avery and Ehrlich 1992, Roach 2002, Field 2005), including the role of L2 speech rhythm. In the case of learners of English as a foreign language (henceforth EFL), non-native rhythm usually results from insufficient durational differences between speech intervals, incorrect word stress and the misapplication of pauses (Adams 1979).

There is still a scarcity of studies on L2 rhythm production, e.g. Guilbault (2002) and Gut (2009) found no effect of pronunciation training or staying abroad on learners' rhythmic properties. A few recent studies (e.g. Lin and Wang 2005, Grenon and White 2008) used different rhythm metrics to compare L2 learners, revealing that their rhythmic scores were located between their L1 and L2 target values. Tortel and Hirst (2010) investigated the rhythmic scores of French learners of English, confirming that rhythm metrics can be successfully used to differentiate between learners and native speakers, as well as learners with varying levels of proficiency, while pointing out the effect of hyperarticulation on the results, a common issue across EFL learners. However, studies on rhythmic differences between learners with different levels of proficiency have been inconclusive. Ordin and Polyanskaya (2015) observed a similar development from syllable-timed patterns to stress-timed speech across L2 learners of English with different L1 rhythm types and various degrees of L2 competence, although German learners of English achieved more native-like rhythm metric scores than corresponding French learners of English, even when the latter group was at an advanced level of proficiency. Jang (2008) found no significant differences between similar groups, while a recent study by White and Mok (2018) reported significant development in L2 rhythm production by observing a change in rhythm metric scores across their subjects, noticing a stronger effect of language experience over length of residence, albeit their sample comprised only five participants.

There is also insufficient research regarding the development of L2 rhythm production among Polish learners of English. Gralińska-Brawata (2014) investigated the effect of pronunciation training on rhythm metric scores in advanced Polish EFL learners. Apparent progress was reported across all participants in at least one rhythm measurement, but only ΔV and VarcoV scores had significant differences between L1 and L2. Although the reliability of rhythm metrics in observing the development of L2 rhythm production still remains inconclusive, the results suggest that language rhythm can be learnable and teachable. Therefore, we expect to notice a difference in the L2 rhythm metric scores before and after training among our participants. Specifically, we expect that the L2 rhythm metric scores will be more similar to native-like results after training due to the acquisition of vowel reduction and greater variation in vowel duration. Moreover, since this study is also interested in the influence of musical aptitude on second language pronunciation, we assume to report more native-like L2 rhythm metric scores before and after training among learners with finer musical aptitude. To summarise, this study will try to answer the following questions:

1. Does accent training affect rhythm metric scores in Polish advanced speakers of English? If yes, which rhythm metrics can indicate this change?
2. Are rhythm metrics capable of differentiating between intermediate and advanced language learners' L2 rhythm production?
3. Does musical aptitude, particularly rhythm perception, affect rhythm metric scores in L2 speech?

2. Method

2.1. Participants

The following study comprised 50 (42 female, 8 male; age 19–21) advanced Polish learners of English (between B2 and C1 according to CEFR framework) with similar results in their secondary school exit exam in English. Before the study, all participants completed LexTALE, i.e. Lexical Test for Advanced Learners of English (Lemhöfer and Broersma 2012), to verify their current level of proficiency in English. The mean result is 74.48% (*SD* = 8.93), which indicates that the group is relatively advanced and homogenous. We observed no L1 accent variation in our participants, i.e. prior to recording English speech, we evaluated impressionistically their Polish spontaneous speech. All participants were using standard Polish (Jassem 2003, Gussmann 2007). Participants had no medically documented speech or hearing impairment and did not have formal accent training prior to the study, i.e. they spoke without a distinctive British or American accent. All participants signed an informed written consent prior to study enrollment and were not remunerated after the study, but received extra course credit for taking part in the study.

2.2. Training

All participants were students of English at university level and received an accent training course in English with a total of 90 hours of class work during the first two semesters as part of their curriculum. The main goal of the course is to teach the students to speak with a native-like General British accent, i.e.

the Standard British English accent spoken in the South of England (students can choose whether to practise General British or General American pronunciation prior to the course start date). The course comprised segmental (i.e. vowels and consonants of English) and suprasegmental phonetics (i.e. intonation, rhythm, syllable stress and sentence stress). The teaching methodology was holistic, i.e. individual features were taught together with prosodic and connected speech processes, rather than in isolation. In the classroom, participants were taught with the use of exercises, reading passages and drama performance. Outside the classroom, participants had to prepare recordings based on various reading texts, dialogues and news reports. During the accent training course, all participants were familiarised with the differences in English vowel length (including such phonetic processes as clipping, i.e. shortening the articulation of vowel before fortis consonants in a stressed syllable) and vowel reduction. These features were practiced in various modes of speech, e.g. via wordlists and reading passages. The participants belonged to four groups and received instruction from four different Polish pronunciation teachers who specialise in teaching English pronunciation and spoke with native-like General British accents. While teaching English pronunciation by non-native speakers can be questionable, the participants should benefit from having instructors with personal experience in learning English pronunciation. Moreover, Polish teachers are aware of typical pronunciation errors made by Polish learners of English and capable of providing sufficient feedback so that students can improve. One of the most common errors made by Polish learners of English is lack of vowel reduction, which is taught during the first academic year and should affect the rhythm metric results (see Sobkowiak 2004 for a comprehensive list). All participants also attended a supplementary practical course in English phonetics and phonology with a total of 45 hours of class work during the first two semesters. The aim of the course is to increase learners' phonological awareness and provide theoretical knowledge for potential future teachers and researchers of English. The course syllabus included articulatory and acoustic phonetics, phonemic and phonetic transcription, intonation and rhythm. Students were assessed via weekly quizzes and four tests.

2.3. Data

A recording of the text "Please Call Stella" was obtained before and after the two-semester accent training course. Both recording sessions were conducted in an anechoic room with an MXL microphone connected to a MacBookPro via a Roland Duo Capture EX audio interface recording in mono 44.1 kHz frequency and 16-bit resolution. The text was displayed on an external monitor in a large white font against a black background. The same procedure was used to record all four pronunciation teachers who conducted the accent training course. The primary reason for analysing L2 English without L1 Polish is related to the fact that native speakers usually have faster speech rate, which in turn can affect the rhythm metric scores, hence the comparison of L2 English and L1 Polish might yield inaccurate results (White and Mattys 2007).

After the first recording session but before the accent training course, all participants completed two musical hearing tests designed by Mandell (2009): the tone deaf test and the rhythm deaf test. The tone deaf test measures pitch perception and melodic memory by playing 36 pairs of synthesised melodies and asking the listener to decide whether each pair is the same or different. The results are expressed in the percentage of correctly identified tokens and use the following scale: below 70% is low, between 70% and 79% is normal, between 80% and 90% is above normal, above 90% is exceptional. The rhythm deaf test measures rhythm perception and rhythmic memory using 25 pairs of synthesised rhythmic patterns and asking the listener to decide whether each pair is the same or different. Similarly to the previous study, the results are expressed in the percentage of correctly identified tokens and use the same scale. Both melodic and rhythmic perception are regarded as good indicators of musical aptitude and both tests are similar in their assumptions and procedures to the ones used in previous studies (Wallentin et al. 2010, Roncaglia-Denissen et al. 2016). Finally, after the musical hearing tests, all participants were asked whether they attended music school or had private music lessons, could play a musical instrument or sing. The answers to these questions were collected via an online survey along with standard demographic data.

After collecting all the data, the recordings were manually edited (i.e. hesitation sounds or repetitions were deleted to avoid their interference in the analysis) and then automatically segmented via FAVE-Extract (Rosenfelder et al. 2014) and Montreal Forced Aligner (McAuliffe et al. 2017) using the DARLA web interface (Reddy and Stanford 2015). Next, all the obtained TextGrid files were verified together with their corresponding audio files and, if necessary, manually corrected using the speech analysis software Praat (Boersma and Weenink 2019). Finally, the rhythm metrics were calculated and plotted on graphs with Correlatore (Mairano and Romano 2010), a programme for rhythmic analysis of annotated Praat TextGrid files. The following rhythm metrics were obtained: Vdev, Cdev, VarcoV, VarcoC, nPVI-V, nPVI-C, CCI-V and CCI-C. All rhythm metrics were computed globally, i.e. including all the values for the consonantal and vocalic intervals in each TextGrid file at once.

3. Results

3.1. Rhythm metrics

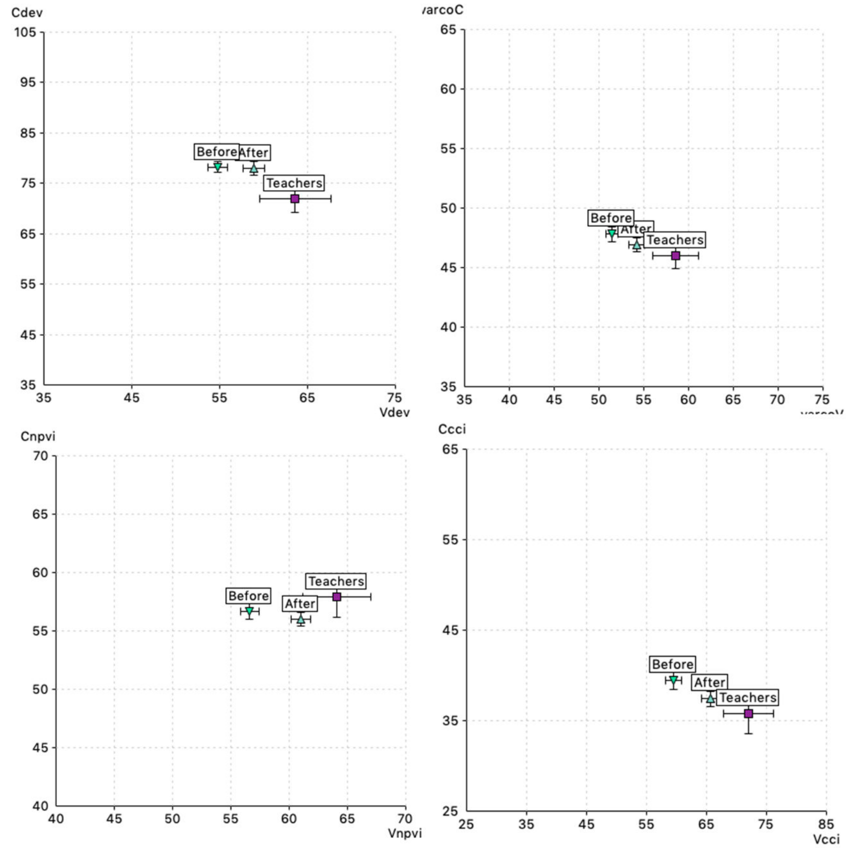

Figure 1. Mean results of participants before and after training with respect to their teachers for individual rhythm metrics: deltas (top left, measuring the standard deviation of V and C intervals), Varco (top right, adding normalisation with respect to speech rate), nPVI (bottom left, measuring the variability of V and C intervals by also considering their succession in time) and CCI (bottom right, measuring the amount of segmental compensation realized by speakers for V and C).

The results presented in Figure 1 show that all rhythm metrics noticed a difference in the participants' rhythmic patterns before and after training. In all cases, the participants' rhythm scores became more similar to their teachers' results after training, taking an intermediate place between the scores before training and the teachers. In terms of language proficiency, this indicates that learners produced more native-like vowel duration contrasts to differentiate between short and long vowels (e.g. short KIT in *bring* and long FLEECE in *please*), as well as produced greater contrasts between stressed and unstressed syllables by using weak forms in their speech (e.g. reducing *to* to /tə/ or *from* to /frəm/). By mastering these characteristic features of English pronunciation, learners' produced more native-like L2 rhythm, since both vowel duration and vowel reduction have a direct impact on English rhythmic patterns.

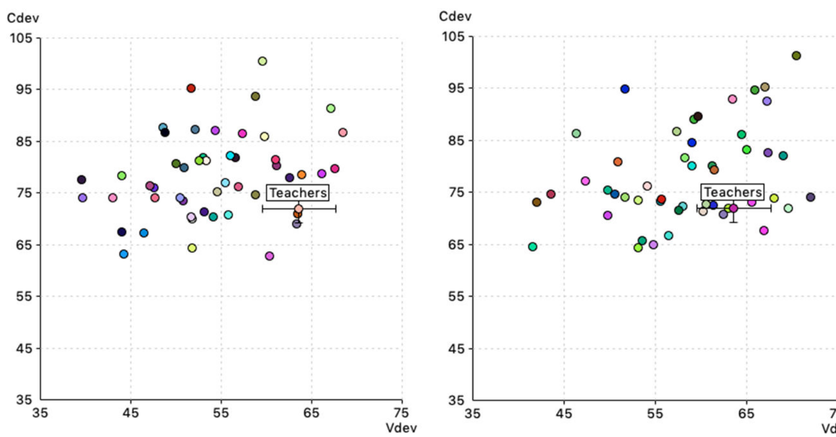

Figure 2. Individual results of participants before (left) and after (right) training with respect to their teachers' mean results for ΔV and ΔC (measuring the standard deviation of V and C intervals).

Figure 2 presents the different ΔV and ΔC scores before and after training. On average, the participants had higher ΔV scores after training ($M = 58.94$, $SE = 1.17$) than before ($M = 54.78$, $SE = 1.15$). Data distribution was normal (before training $W(50) = 0.97$, $p = .23$, after training $W(50) = 0.98$, $p = .79$). A dependent one-tailed *t*-test was performed to compare the rhythm metric scores before and after training. The difference was significant ($t(49) = 3.34$; $p < .01$). However, there was no observable change in the ΔC score before ($M = 78.12$, $SE = 1.15$) and after training ($M = 78.02$, $SE = 1.27$). These results suggest that participants generally improved in vowel reduction and produced more variable vowel duration, while their consonantal variation remained unchanged.

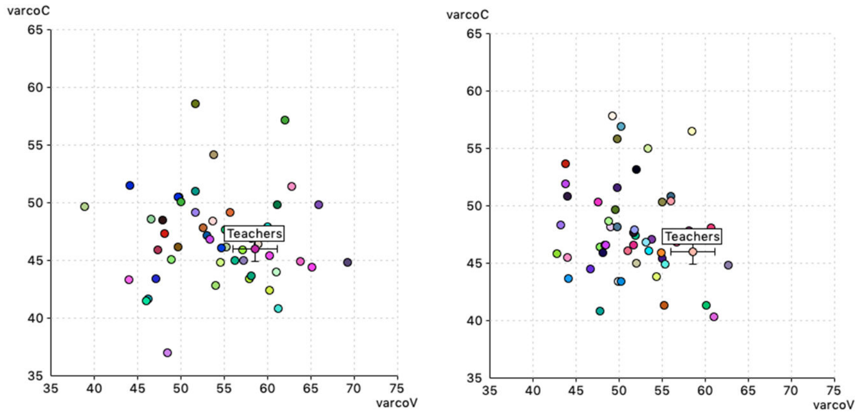

Figure 3. Individual results of participants before (left) and after (right) training with respect to their teachers' mean results for VarcoV and VarcoC (measuring the standard deviation of V and C intervals while adding normalisation with respect to speech rate).

Figure 3 presents the different VarcoV and VarcoC scores before and after training. Participants scored higher in VarcoV after training (M = 54.34, SE = 0.91) than before (M = 51.48, SE = 0.70). Data distribution was normal (before training $W(50) = 0.96$, $p = .16$, after training $W(50) = 0.99$, $p = .95$). The difference between the results was significant ($t(49) = 3.38$; $p < .01$). Similarly to ΔC, there was no observable difference for VarcoC before (M = 47.83, SE = 0.59) and after training (M = 46.93, SE = 0.56). Overall, we can observe more dispersed results before training than after training, while the results after training are closer to the teachers' results.

The nPVI results presented in Figure 4 show a similar case to the previous rhythm metrics, with more scattered vocalic results before training (M = 56.61, SE = 0.76), while the same metrics after training are aligned closer to the teachers' scores (M = 60.96, SE = 0.77). Data distribution was normal (before training $W(50) = 0.96$, $p = .11$, after training $W(50) = 0.98$, $p = .71$). The difference between the results was significant ($t(49) = 5.21$; $p < .01$). Again, there was no observable difference before (M = 56.73, SE = 0.72) and after (M = 56.04, SE = 0.59) training for the consonantal nPVI.

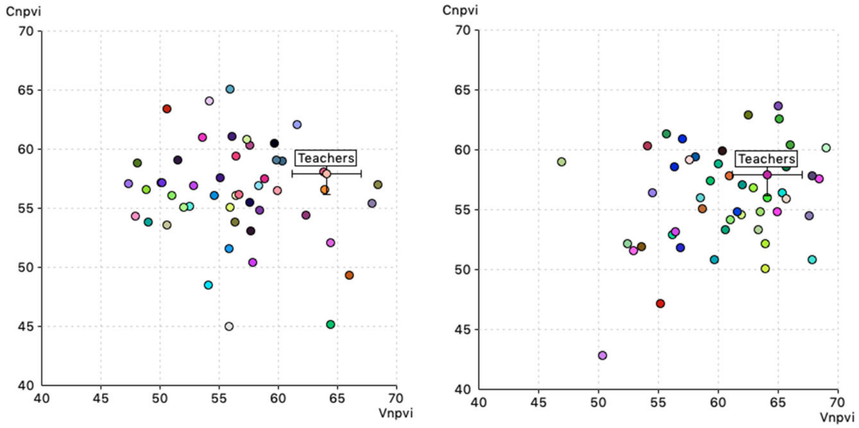

Figure 4. Individual results of participants before (left) and after (right) training with respect to their teachers' mean results for nPVI-V and nPVI-C (measuring the variability of V and C intervals by also considering their succession in time).

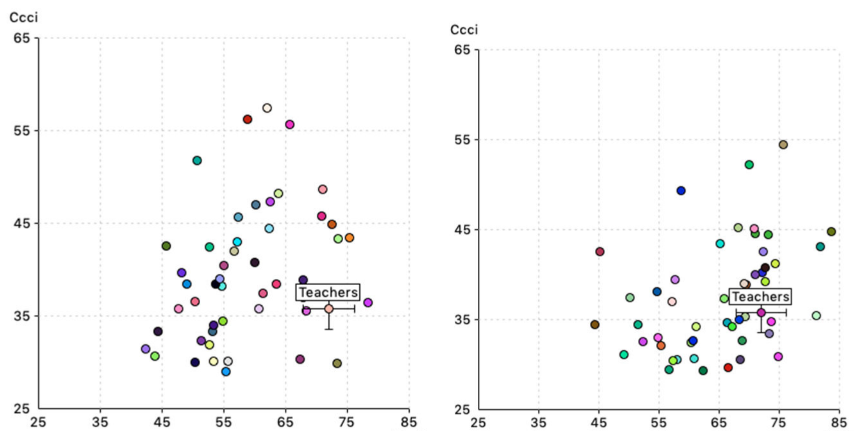

Figure 5. Individual results of participants before (left) and after (right) training with respect to their teachers' mean results for CCI-V and CCI-C (measuring the amount of segmental compensation realised by speakers for V and C).

Finally, the CCI results in Figure 5 show a similar pattern to the previous rhythm metrics, where the vocalic scores after training (*M* = 65.55, *SE* = 1.42) are closer to the teachers' scores than before (M = 59.50, SE = 1.45). Data distribution was normal (before training $W(50) = 0.96$, $p = .06$, after training $W(50) = 0.97$, $p = .39$). The difference between the results before and after training was significant ($t(49) = 4.13$; $p < .01$). Interestingly, we also reported a significant difference for the consonantal scores before (M = 39.53, SE = 1.02) and after (M = 37.41, SE = 0.85) training ($t(49) = -2.44$; $p < .01$). Since the CCI is designed to measure the amount of compression at a segmental level, languages considered as syllable-timed are expected to have comparable variability of vocalic and consonantal segments, while traditionally stress-timed languages have high variability of vocalic segments and low variability of consonantal segments. Therefore, the result for CCI-C suggests that the participants managed to produce more native-like L2 consonant clusters after training with lower consonantal variation, similar to their teachers. This is also in line with previous studies incorporating the CCI metric (Mairano and Romano 2011).

To summarise, all rhythm metrics allowed us to observe a significant progress in the vocalic scores after training, suggesting that accent training can help students in achieving more native-like vowel duration and vowel reduction, which has a direct effect on perceived speech rhythm in English. At the same time, most rhythm metrics observed no significant change in the consonantal scores after training. This can be related to the fact that the difference in the rhythm metric scores for Polish and English consonantal intervals is much smaller than for vocalic intervals (Mairano and Romano 2011: 1320-1321), leading Polish learners of English to focus on acquiring L2 English vowel duration contrasts and mastering vowel reduction. In conclusion, the results confirm that the above-mentioned rhythm metrics can be successfully used to study a change in L2 rhythm production, even in case of upper-intermediate and advanced learners of English after a year of accent training.

3.2. Musical hearing tests and survey results

Figure 6 presents the tone deaf and rhythm deaf test results, which express the percentage of correctly identified tokens. The mean result for the tone deaf test was 68% (min = 44.4%, max = 86.1%, median 69.4%), while the mean result for the rhythm deaf test was 71.2% (min = 48%, max = 92%, median 72%). Both test results had normal distribution (tone deaf $W(50) = 0.96$, $p = .13$, rhythm deaf $W(50) = 0.95$, $p = .054$). By comparing the results with the provided scale, we can say that in the tone deaf test 27 participants scored below

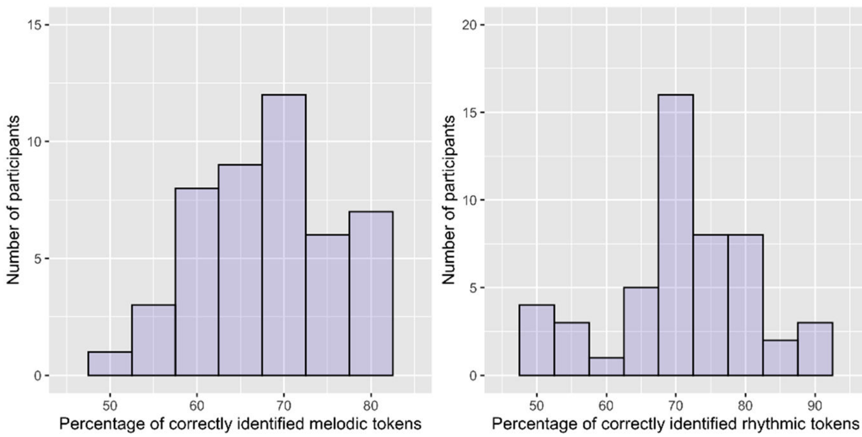

Figure 6. Tone deaf (left) and rhythm deaf (right) test results.

normal (< 70%), 17 normal (70%–79%), and six above normal (80%–90%), while in the rhythm deaf test 20 participants scored below normal (< 70%), 17 normal (70%–79%), 11 above normal (80%–90%), and two exceptional (> 90%).

In the survey, four participants claimed to have completed music school, 12 participants have played a musical instrument, and 12 have sung as soloists, band members or in a choir.

3.3. Musical aptitude and L2 rhythm production

To verify whether musical aptitude is related to L2 rhythm production, we compared the musical hearing test results with the vocalic scores obtained from the rhythm metrics after training. Pearson's Correlation Coefficient reported none of the correlations between the melodic deaf test scores or the rhythm deaf test scores and the individual rhythm metric results after training as significant. All r and p results are presented in Table 1.

To see if the musical hearing tests correlate with the closeness to the model pronunciation, we also calculated the percentage differences between the participants' rhythm metric scores for vocalic intervals with their teachers' scores and compared them with the participants' musical hearing scores using Pearson's Correlation Coefficient. We observed a weak positive correlation be-

Table 1. Correlation between melodic deaf test (MDT) and rhythm deaf test (RDT) scores and rhythm metric scores before (B) and after (A) training.

		ΔV		VarcoV		nPVI-V		CCI-V	
		B	A	B	A	B	A	B	A
MDT	*r*	−0.08	−0.05	−0.006	0.07	−0.005	0.07	−0.14	−0.10
	p	.58	.73	.96	.62	.97	.62	.33	.48
RDT	*r*	0.15	0.21	0.11	-0.02	0.11	-0.13	0.08	0.13
	p	.28	.14	.44	.89	.44	.36	.54	.36

tween the varcoV scores and the rhythm deaf test results ($r = 0.21$, $df = 49$, $p = .13$), as well as between the nPVI-V scores and the rhythm deaf test results ($r = 0.16$, $df = 49$, $p = .26$). To investigate further, we divided the participants into two groups based on their musical hearing tests (80% being the cut-off point between the below-average and above-average) but found no significant differences in varcoV scores between groups. We also divided the participants into two groups: "musicians", i.e. participants who confirmed in the survey to have completed music school (4) and/or have played a musical instrument (12) and/or have sung as soloists/in a band/in a choir (12), and "non-musicians", i.e. participants without any specific musical experience (33). The result was insignificant for all rhythmic scores after training. These results suggest that musical aptitude or musical experience might not play an important role in the acquisition of L2 rhythm during a one-year accent training course in a formal academic learning environment. However, it is possible that the influence of musical hearing and musical background on the production of L2 rhythm might not be visible due to the method used for assessing L2 rhythm, the relatively short duration of the accent training course, or its formal academic context.

4. Discussion

The acquisition of L2 speech rhythm is a complex issue that can be difficult to measure and is affected by numerous linguistic and extra-linguistic factors.

Our findings suggest that rhythm metrics can be successfully used to observe improvement in L2 rhythm production in a longitudinal study, particularly in regards to increased vocalic variation in L2 English. Interestingly, all rhythm metrics yielded similar results, showing the mean progress of our participants after training and occupying an intermediate place between their rhythm metric scores before training and the mean rhythm metric scores of their pronunciation teachers. Furthermore, the results show that L2 rhythm is teachable and learnable in a formal setting during an accent training course. In this particular case, we observed that Polish adult advanced learners of English changed the degree of their vocalic variation after one year of training, indicating successful acquisition of variable vowel duration and vowel reduction. Finally, we did not find any significant relationship between the musical aptitude results and the rhythm metric scores.

When discussing rhythm metrics, it is vital to remember that these measurements rely on durational properties of speech and do not incorporate other phonological, lexical and syntactic factors affecting perceived speech rhythm (Dauer 1983; Gut 2012). In this regard, language rhythm presented with the use of these metrics is a combination of vocalic and consonantal variation, which are phonetic outcomes of various phonological processes. Therefore, instead of teaching L2 rhythm in a conscious and formal manner with the use of metalanguage, Barry (2007) suggests to concentrate on the underlying phonological processes that affect the perceived speech rhythm in a given language. In the case of EFL pronunciation teaching, the primary focus should be placed on vowel duration and vowel reduction, which have a major effect on English rhythm and are also specifically related to rhythm metric scores. This was also the main method of teaching pronunciation in the accent training course in this study, as pronunciation teachers focused on segmental and suprasegmental phonetics by presenting them in a practical and holistic approach in the classroom, while formal context and metalanguage was discussed in the supplementary course on English phonetics and phonology.

The reason why we did not observe any significant relation between musical hearing tests and L2 rhythm production can be twofold. First, the apparent progress made by our participants might be primarily caused by the accent training course alone, suggesting that all participants can equally benefit from formal instruction when acquiring L2 pronunciation, regardless of their musical aptitude or musical experience. An important factor that should also be addressed is motivation, which can be a valid predictor in successful acquisition of L2 pronunciation among advanced adult learners of English (Smit 2002). Although we tried to control for participants' general proficiency, as well as their performance in and outside the classroom, it is possible that some learners spent more private time on practicing pronunciation. As this was a longitudinal study comparing results before and after one year of studies, individual motivation of our participants could affect the amount of effort they put in during the accent training course, thus, participants with higher motivation could compensate for their lower musical hearing test scores. To verify this in future longitudinal studies, it is crucial to take into account participants' motivation and control for their work outside the classroom. Finally, it is possible that musical hearing can be related to aspects of language rhythm not measured by rhythm metrics, such as f0 or intensity, which also affect perceived speech rhythm (Cumming 2011). Although there is a reported difference between the processing of pitch in language and music (Zatorre and Baum 2012), studies suggest that musical training can improve pitch processing (related to F0) in both domains (Schön et al. 2004). Therefore, it is possible that musical hearing will not directly correlate with rhythm metrics scores, but can still have an effect on successful L2 rhythm acquisition.

The following study has two limitations that should be taken into consideration in future research. First, while the use of rhythm metrics to observe progress across advanced EFL learners proved to be successful, it is crucial to remember that these metrics focus only on the segmental distribution and durational contrasts, omitting other features that can influence language rhythm, such as pitch (Pickering and Wiltshire 2000), sonority (Galves et al. 2002), and loudness (Fuchs 2014). Therefore, it would be valuable to consider adding these features in future studies. Alternatively, it would be beneficial to implement phonological models of rhythm (e.g. Hayes 1995) focusing on prominence patterns and shift towards a more psychological understanding of language rhythm (Arvaniti 2009). Second, while the use of musical hearing tests by Mandell (2009) was a novel approach, it would be also valuable to include tests evaluating music production (Wallentin et al. 2010), as the assessment of actual musical performance (i.e. singing or playing a musical instrument) could provide more authoritative results than participants' self-reported years of musical training. Furthermore, it would be also advantageous to implement tests specifically designed to evaluate participants' rhythmic skills (Bella et al. 2017) and correlate these results with their potential progress in L2 rhythm production. All in all, incorporating these extensions should provide a more complete picture of language rhythm and its potential relation to musical aptitude.

To conclude, this study confirmed the usefulness of rhythm metrics in reporting progress in the acquisition of L2 rhythm in Polish advanced learners of English, complementing earlier studies in this field (e.g. Gralińska-Brawata 2014). At the same time, we did not find any relationship between L2 rhythm metric scores and musical hearing or musical experience. Possibly, musical skills and music education could be related to other factors that constitute speech rhythm and are not expressed by rhythm metrics. Therefore, it is vital to include other factors that constitute L2 speech rhythm, find new ways to quantify language rhythm, as well as include other forms to assess participants' musical hearing and their musical background. We hope that this research will direct future quantitative and longitudinal studies in the field of L2 rhythm production and the role of musical hearing in L2 pronunciation.

5. Acknowledgments

This research was supported by the National Science Centre in Poland, grant number 2014/15/N/HS2/03865.

REFERENCES

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

Adams, C. 1979. *English Speech Rhythm and the Foreign Learner*. The Hague: Moulton.

Arvaniti. A. 2009. "Rhythm, timing and the timing of rhythm". *Phonetica* 66(1-2). 46–63.

Arvaniti, A. 2012. "The usefulness of metrics in the quantification of speech rhythm". *Journal of Phonetics*, 40(3). 351–373.

Avery, P. and S. Ehrlich. 1992. *Teaching American English Pronunciation*. Oxford: Oxford University Press.

Barry, W. J. 2007. "Rhythm as an L2 problem. How prosodic is it?". In: Trouvain, J. and U. Gut (eds.) *Non-native Prosody: Bridging the Gap between Research and Teaching*. 97–120.

Bella, D., Farrugia, N., Benoit, C., Begel, V., Verga, L., Harding, E., Kotz, S. 2017. "BAASTA: Battery for the Assessment of Auditory Sensorimotor and Timing Abilities". *Behavior Research Methods* 49. 1128–1145.

Bertinetto, P.M. and C. Bertini. 2008. On modeling the rhythm of natural languages. *Proceedings of Speech Prosody 2008*, Campinas, Brazil. 427–430.

Boersma, P. and D. Weenink. 2019. Praat: doing phonetics by computer [Computer program]. Version 6.1.03, retrieved 1 September 2019 from http://www.praat.org/

Chobert, J. and M. Besson. 2013. "Musical Expertise and Second Language Learning". *Brain Sciences,* 3(2). 923–940.

Cumming R.E. (2011). "The language-specific interdependence of tonal and durational cues in perceived rhythmicality". *Phonetica* 68. 1–25.

Llanes-Coromina, J., P. Prieto and P.L. Rohrer. 2018. "Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task.". *9th International Conference on Speech Prosody 2018*. 498–502.

Coumel, M., M. Christiner and S.M. Reiterer. 2019. "Second Language Accent Faking Ability Depends on Musical Abilities, Not on Working Memory". *Frontiers in psychology* 10(257).

Culp, M.E. 2017. "The Relationship Between Phonological Awareness and Music Aptitude". *Journal of Research in Music Education* 65(3). 328–346.

Dauer R.M. 1983. "Stress-timing and syllable-timing reanalyzed". *Journal of Phonetics* 11. 51–62.

Dellwo, V. and P. Wagner. 2003. "Relations between language rhythm and speech rate". In: Solé, M., D. Recasens and J. Romero (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences*, 3–9 August, Barcelona, Spain. 471–474.

Dellwo, V. 2006. "Rhythm and speech rate: A variation coefficient for deltaC". In: Karnowski, P. and I. Szigeti (eds.). *Language and language processing: Proceedings of the 38th linguistic Colloquium*. Piliscsaba 2003. Frankfurt: Peter Lang. 231–241.

Fabricius, A., D. Watt and D. Johnson. 2009. "A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics". *Language Variation and Change* 21(3). 413–435.

Fadiga L., L. Craighero and A. D'Ausillo. 2009. "Broca's area in language, action, and music". *Annals of the New York Academy of Sciences* 1169. 44–458.

Fenk-Oczlon, G. and A. Fenk. 2009. "Some parallels between language and music from a cognitive and evolutionary perspective". *Musicae Scientiae* 13(2). 201–226.

Field, J. 2005. "Intelligibility and the Listener: the Role of Lexical Stress". *TESOL Quarterly* 39. 399–423.

Franklin, M., K. Moore, C. Yip and J. Jonides. 2008. "The effects of musical training on verbal memory". *Psychology of Music* 36. 353–365.

Fuchs, R. 2019. "Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English". *Speech Prosody* 7. 290–294.

Gabriel, B. 2007. *Learning English Through Songs*. Singapore: Bettyland Publications.

Galves, A., J. Garcia, D. Duarte and C. Galves. 2002. "Sonority as a basis for rhythmic class discrimination". *Proceedings of Speech Prosody 2002*. 323–326.

Gibbon, D. 2003. "Computational modelling of rhythm as alternation, iteration and hierarchy". In: Solé, M., D. Recasens and J. Romero (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences*, 3–9 August, Barcelona, Spain. 2489–2492.

Gordon, E.E. 1989. *Learning sequences in music: Skill, content, and patterns*. Chicago: G. I. A. Publications.

Grabe, E. and E. Low. 2002. "Durational variability in speech and the rhythm class hypothesis". *Laboratory Phonology* 7. 515–546.
Gralińska-Brawata, A. 2014. "Language Experience and Phonetic Training as Factors Influencing Timing Organisation in Polish Learners of English". *Research in Language* 12. 185–198.
Grenon, I. and L. White. 2008. "Acquiring Rhythm: a Comparison of Canadian English and Japanese". In: Chan, H., H. Jacob and E. Kapia (eds.), *BUCLD, Proceedings of the 32nd annual Boston University Conference on Language development*. Somerville, Massachusetts: Cascadilla Press 32. 155–166.
Guilbault, C. 2002. The Acquisition of French Rhythm by Second Language Learners. PhD thesis, University of Alberta.
Gussmann, E. 2007. *The Phonology of Polish*. Oxford: Oxford University Press.
Gut, U. 2009. *Non-native Speech: a Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt: Peter Lang.
Gut, U. 2012. "Rhythm in L2 speech". In: Gibbon, D., D. Hirst and N. Campbell (eds.), *Rhythm, Melody and Harmony in Speech: Studies in Honour of Wiktor Jassem, Special edition of Speech and Language Technology* 14/15. Poznań: Polskie Towarzystwo Fonetyczne. 83–94.
Hayes, B. 1995. *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.
Jang, T.Y. 2008. "Speech Rhythm Metrics for Automatic Scoring of English Speech by Korean EFL Learners". *Malsori Speech Sounds. The Korean Society of Phonetic Sciences and Speech Technology* 66. 41–59.
Jassem, W. 2003. "Polish". *Journal of the International Phonetic Association* 33(1), 103–107. https://doi.org/10.1017/S0025100303001191
Jusczyk, P.W. 1999. "Narrowing the distance to language: one step at a time". *Journal of Communication Disorders* 32. 207–222.
Koelsch, S., T.C. Gunter, D.Y. von Cramon, S. Zysset, G. Lohmann and A.D. Friederici. 2002. "Bach speaks: a cortical 'language-network' serves the processing of music". *Neuroimage* 17. 956–966.
Lemhöfer, K. and M. Broersma. 2012. "Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English". *Behavior Research Methods* 44. 325–343.
Lin, H. and Q. Wang. 2005. "Vowel quantity and consonant variance: A comparison between Chinese and English". Paper presented at the Conference Between Stress and Tone, Leiden.
London, J. and K. Jones. 2011. "Rhythmic Refinements to the nPVI Measure: A Reanalysis of Patel & Daniele (2003a)". *Music Perception* 29. 115–120.
Low, E.L., E. Grabe and F. Nolan 2000. "Quantitative characterization of speech rhythm: Syllable-timing in Singapore English". *Language and Speech* 43. 377–401.
Magne, C., D. Schön and M. Besson. 2006. "Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches". *Journal of Cognitive Neuroscience* 18. 199–211.

Magne, C., D.K. Jordan and R.L. Gordon. 2016. “Speech rhythm sensitivity and musical aptitude: ERPs and individual differences”. *Brain and Language* 153/154. 13–19.

Mandell, J. 2009. Electronic Music and Medical Education. (http://jakemandell.com) (date of access: 21 September 2019).

Mairano, P. and A. Romano. 2010. Un confronto tra diverse metriche ritmiche usando Correlatore. In: Schmid, S., Schwarzenbach, M. and D. Studer (eds.). *La dimensione temporale del parlato*, (Proc. of the V National AISV Congress, University of Zurich, Collegiengebaude, 4-6 February 2009), Torriana (RN): EDK, 79–100.

Mairano P. and A. Romano. 2011. “Rhythm metrics for 21 languages”. In: *Proceedings. of the 17th International Congress of Phonetic Sciences*, Hong Kong, China. 1318–1321.

McAuliffe, M., M. Socolof, S. Mihuc, M. Wagner and M. Sonderegger. 2017. “Montreal Forced Aligner: trainable text-speech alignment using Kaldi”. In: *Proceedings of the 18th Conference of the International Speech Communication Association*.

Milovanov, R., M. Tervaniemi and M. Gustafsson. 2004. “The impact of musical aptitude in foreign language acquisition”. In: *Proceedings of the 8th International Conference on Music Perception and Cognition*.

Milovanov, R., M. Huotilainen, V. Välimäki, P.A. Esquef and M. Tervaniemi. 2008. “Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence”. *Brain Research* 1194. 81–89.

Milovanov, R., P. Pietilä, M. Tervaniemi and P.A. Esquef. 2010. “Foreign language pronunciation skills and musical aptitude: a study of Finnish adults with higher education”. *Learning and Individual Differences* 20. 56–60.

Moritz, C., S. Yampolsky, G. Papadelis, J. Thomson and M. Wolf. 2013. “Links between early rhythm skills, musical training, and phonological awareness”. *Reading and Writing: An Interdisciplinary Journal* 26(5). 739–769.

Nazzi T., J. Bertoncini and J. Mehler. 1998. “Language discrimination by newborns: toward an understanding of the role of rhythm”. *Journal of Experimental Psychology: Human Perception and Performance* 24. 756–766.

Nespor, M. 1990. “On the rhythm parameter in phonology”. In: Roca, I. M. (ed.). *Logical Issues in Language Acquisition*, Dordrecht: Foris: 157–175.

Nettl, B. 2000. “An ethnomusicologist contemplates universals in musical sound and musical culture”. In: Wallin, N.L., B. Merker and S. Brown (eds.), *The Origins of Music*. Cambridge. MA: MIT Press. 463–472.

Ordin, M. and L. Polyanskaya. 2015. “Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2”. *Frontiers in Psychology* 6. 1–15.

Pastuszek-Lipińska, B. 2008. "Musicians outperform nonmusicians in speech imitation". *Lecture Notes in Computer Science* 4969. 56–73.

Patel, A. and J.R. Daniele. 2003. “An empirical comparison of rhythm in language and music”. *Cognition* 87(1). 35–45.

Patel, A. 2008. *Music, Language, and the Brain*. New York: Oxford University Press.

Pickering, L. and C. Wiltshire. 2000. “Pitch accent in indian-english teaching discourse”. *World Englishes* 19(2). 173–183.

Pike, K.L. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.

Ramus F., M. Nespor and J. Mehler. 1999. "Correlates of linguistic rhythm in the speech signal". *Cognition* 73. 265–292.

Ramus, F. 2002. "Acoustic correlates of linguistic rhythm: Perspectives". *Proceedings of Speech Prosody*. 115–120.

Ramus, F., E. Dupoux and J. Mehler. 2003. "The psychological reality of rhythm classes: Perceptual studies". In: Solé, M., D. Recasens and J. Romero (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences*, 3–9 August, Barcelona, Spain. 337–342.

Reddy, S. and J. Stanford. 2015. "A Web Application for Automated Dialect Analysis". *Proceedings of NAACL-HLT 2015*.

Roach, P. 1982. "On the distinction between 'stress-timed' and 'syllable-timed' languages", in: Crystal, D. (ed.). *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour of F.R. Palmer*. London. 73–79.

Roach, P. 2002. "Studying rhythm and timing in English speech: Scientific curiosity, or a classroom necessity?". In: Waniek-Klimczak, E. and P. J. Melia (eds.). *Accents and Speech in Teaching English Phonetics and Phonology. Łódź Studies in Language Vol.5*. Frankfurt am Main: Peter Lang. 199–206.

Roncaglia-Denissen, M.P, D.A. Roor, A. Chen and M. Sadakata. 2016. "The Enhanced Musical Rhythmic Perception in Second Language Learners". *Frontiers in Human Neuroscience* 10.

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

Rubach, J. and G.E. Booij, 1985. "A grid theory of stress in Polish". *Lingua* 66. 281–319.

Schön, D., Magne, C., and M. Besson. 2004. "The music of speech: Music training facilitates pitch processing in both music and language". *Psychophysiology* 41(3). 341–349.

Slevc, L. and A. Miyake. 2006. "Individual differences in second language proficiency: Does musical ability matter?". *Psychological Science* 17. 675–681.

Smit, U. 2006. "The interaction of motivation and achievement in advanced EFL pronunciation learners". *IRAL - International Review of Applied Linguistics in Language Teaching* 40(2). 89–116.

Sobkowiak, W. 2004. *English phonetics for Poles*. (3rd edition.) Poznań: Wydawnictwo Poznańskie.

Tortel, A. and D. Hirst. 2010. "Rhythm metrics and the production of English L1/L2". *Speech Prosody 2010*.

Wallentin, M., A.H. Nielsen, M. Friis-Olivarius, C. Vuust and P. Vuust. 2010. "The musical ear test, a new reliable test for measuring musical competence". *Learning and Individual Differences*. 20. 188–196.

White L. and S.L. Mattys. 2007. "Calibrating rhythm: first language and second language studies". *Journal of Phonetics* 35. 501–522.

White D. and P. Mok. 2018. "L2 Speech Rhythm Development in New Immigrants". *9th International Conference on Speech Prosody 2018*. 838–842.

Wong, P., E. Skoe, N. Russo, T. Dees and N. Kraus. 2007. "Musical experience shapes human brainstem encoding of linguistic pitch patterns". *Nature Neuroscience* 10. 420–422.
Zatorre, R. J., and S.R. Baum. 2012). "Musical Melody and Speech Intonation: Singing a Different Tune". *PLoS Biology* 10(7). e1001372.

Address correspondence to:
Mateusz Jekiel
Faculty of English
Adam Mickiewicz University
Collegium Heliodori Święcicki
Grunwaldzka 6
60-780 Poznań
mjekiel@amu.edu.pl