The Influence of the Mother Tongue and of Musical Experience on Rhythm Perception

Sumio Kobayashi

A dissertation submitted to the University of Kent for the degree of Doctor of Philosophy.

School of European Culture and Languages

31st January 2020

Abstract

Native language and musical experience are both said to influence our perception of rhythm; however, the study of the influence of native language on rhythm perception is limited. This thesis tested if and how linguistic and musical experiences affect our rhythm perception. The term rhythm, as used here, is identical to the musical term, metre, which refers to a recurring regular pattern of prominent and non-prominent elements. First, this thesis examined language-specific rhythms in English, Japanese, and Russian to explore whether listeners are better at detecting irregularities in rhythms that frequently occur in their native language, compared to those that are less frequent. A review of the existing literature and an original, corpus-based examination show that English and Russian rhythms are based on a relatively regular alternation of prominent and non-prominent syllables, whereas Japanese rhythm is based on a subtle alternation of prominent and non-prominent morae, less regular than that of English and Russian rhythms. Similarly, culture-specific musical rhythms are discussed to examine the influence of musical experience on rhythm perception. It is shown that, in traditional Japanese and Russian musical works, non-binary rhythms are prevalent, while they are relatively rare in English music. A series of perceptual experiments with both English, Japanese, and Russian-speaking musicians and non-musicians showed that musical experience affects rhythm perception but is less effective than linguistic experience in shaping responses to rhythm irregularities. These perception experiments showed that Japanese speakers perceived binary and non-binary rhythms more accurately than English and Russian speakers, while there were no significant differences between English and Russian speakers. In addition, it was found that clashes (rhythm irregularities caused by successive prominent elements) were less tolerated than lapses (rhythm irregularities caused by sequences of non-prominent elements). The experimental results showed that all participants tolerated lapses more readily than clashes, which suggests that clashes lead to dysrhythmic sequences that are easier to detect than those of lapses.

Contents

Chapter 1: Introduction

Chapter 2: English, Japanese and Russian rhythms

Chapter 3: Musical Rhythm

Chapter 4: Corpus-based study of Japanese rhythm

Chapter 5: Methods of psychological experiments

Chapter 6: The role of linguistic experience: English vs. Japanese vs. Russian

Chapter 7: The role of musical experience: English and Japanese musicians and non-musicians

Chapter 8: General discussion on the influence of linguistic and musical experience and conclusion

List of tables

Chapter 2: English, Japanese and Russian rhythms

Chapter 3: Musical Rhythm

Chapter 4: Corpus-based study of Japanese rhythm

Chapter 5: Experimental methods

Chapter 6: The role of linguistic experience: English vs. Japanese vs. Russian

Chapter 7: The role of musical experience: English vs. Japanese musicians and non-musicians

List of figures

Chapter 1: Introduction

Chapter 3: Musical Rhythm

Chapter 4: Corpus-based study of Japanese rhythm

Chapter 5: Experimental methods

Chapter 6: The role of linguistic experience: English vs. Japanese vs. Russian

Chapter 7: The role of musical experience: English and Japanese musicians and non-musicians

Chapter 1: Introduction

1.1 Preliminary remarks

This thesis deals with the perception of rhythm and addresses three main questions:

(i) How is rhythm perception influenced by the rhythm of one's native language?

(ii) How is rhythm perception influenced by one's musical experience, both in terms of musical training and exposure to different types of rhythm?

(iii) How are different types of stimuli (linguistic, musical, and tonal) processed in terms of rhythm?

In looking at past studies, examination of the research questions above is of interest for the following reasons:

(a) as detailed later in Section 3.1, there is disagreement among researchers about what rhythm is and whether it is the same in both music and language;

(b) some studies show that there is a connection between native language rhythm and musical rhythm within a culture (e.g. Patel, 2003), but studies such as Patel (2003) suggest there is a connection between native language rhythm and musical rhythm within a culture, but have only used data from English and French. This thesis extends this line of research to include Japanese and Russian, which exhibit different characteristics to English and French, both in terms of linguistic rhythm and musical tradition;

(c) the above means that we do not know if listeners react in the same way to musical and linguistic stimuli. To examine this from various perspectives, tonal stimuli are included, as many studies on rhythm perception have focused either on the relationship between language and music or derive from the literature on the psychology of rhythm and rely on tonal stimuli.

Providing answers to these questions will help with understanding rhythm perception in general, the relationship between speech and music, and the role of linguistic and musical experiences in the processing of linguistic prosody.

To address these questions, I tested groups of listeners from different native languages – English, Japanese, and Russian – and distinct musical traditions that rely on different types of rhythm. Here, I adopt a psychological definition of rhythm that applies to music as well: rhythm will be taken to refer to the regular alternation of prominent and non-prominent elements that create groups of alternating patterning (further discussion can be found in Arvaniti, 2009). A more detailed definition of rhythm is given in Section 1.3. More succinctly, this study concerns rhythm perception, defining rhythm as a binary pattern of accented and unaccented elements (stress-accented syllables in English and Russian, pitch-accented mora in Japanese, and accented notes in music). When examining how native language affects rhythm perception, past studies on the influence of native language on perception are useful to consider, including studies on the perception of consonants and vowels.

A number of studies show that native language affects how we perceive and categorise consonants and vowels. For instance, Goto (1971) shows that Japanese speakers cannot distinguish English /ɹ/ (alveolar approximant) from /l/ (alveolar lateral approximant). Logan, Lively, and Pisoni (1991) suggest that the difficulty of discriminating /r/ and /l/ shown in Goto (1971) can be eased through training, but not perfectly. The results in Goto (1971) have been replicated on many occasions since then, for example, in Aoyama, Flege, Guion, Akahane-Yamada, and Yamada (2004); Bradlow, Akahane-Yamada, Pisoni, and Tohkura (1999); Flege, Takagi, and Mann (1995); Lively, Logan, and Pisoni (1993); and Logan, Lively, and Pisoni (1991). Studies such as Beddor and Strange (1982) and Bohn and Flege (1990) show that vocalic perception is influenced by the native language.

In addition to segment perception, native language affects the perception of prosodic phenomenon; for instance, Cutler (2000) showed that stress accent is a cue English speakers use to locate word boundaries, not only when listening to English, but also when listening to other languages (see Section 2.1.1 about stress in English). These findings suggest that rhythm perception can also be influenced by native language.

Moreover, several studies indicate that our non-linguistic sound perception differs depending on the mother tongue. The relationship between linguistic experience and non-linguistic sound perception has been examined in a number of studies (e.g. Bent, Bradlow, & Wright, 2006; Deutsch, 1991; Deutsch, Henthorn, & Dolson, 2004; Deutsch, Henthorn, Marvin, & Xu, 2006;).

Bent et al. (2006), Deutsch (1991), Deutsch et al. (2004), and Deutsch et al. (2006) all examined whether native language affects non-linguistic pitch perception, by examining the demography of holding absolute pitch (an ability to detect pitch precisely) among speakers of tone languages (Mandarin or Vietnamese) and non-tonal language (English) speakers. Their results showed that tone language speakers are more likely to have absolute pitch than those of non-tonal languages.

Grouping studies examined whether native language affects how listeners group a series of stimuli – a task that relates to the perception of rhythm (e.g. Bion, Benavides-Varela, & Nespor, 2011; Iversen, Patel, & Ohgushi, 2008; Jeon & Arvaniti, 2016; Kusumoto & Moreton, 1997; Yoshida et al., 2010). These studies dealt with whether listeners showed a preference for iambs (prominent element last, i.e. weak-strong) or trochees (prominent element first, i.e. strong-weak) in listening stimuli. Iversen et al. (2008) examined whether the perception of sound grouping is different depending on the mother tongue, by comparing native speakers of Japanese and American English. They showed that, when listening to a series of alternating long and short tones, Japanese speakers tended to perceive them as trochees (long-short), while English speakers perceived them as iambs (short-long). On the other hand, Iversen et al. (2008) show that 90% of Japanese and 63% of English speakers perceived trochees when listening to a series of alternating loud-soft tones. Iversen et al. (2008) argue that Japanese speakers tend to perceive sequences of alternating long and short tones as trochees, because Japanese rhythm tends to be trochaic;

for example, a phrase such as '**U**ga **to**ru **te**wa **yo**i' (the way which a cormorant chooses is good). [ʻUgaʻtoruʻtewaʻjoi] is a trochaic Japanese phrase, due to the fact that function words follow content words and are not accented. They interpreted this as evidence that function words act as weak elements in rhythm. Furthermore, they argued that English speakers tend to perceive stimuli alternating in duration as having iambic rhythm (weak-strong pattern), because English rhythm favours iambic as function words, which are weak elements, precede content words; for example, 'the **sun** is **hot'** is an iambic phrase with the unstressed function words 'the' and 'is', preceding the stressed words 'sun' and 'hot'.

Similarly, Bion, Benavides-Varela and Nespor (2011), Kusumoto and Moreton (1997), and Yoshida et al (2010) examined whether perceptual grouping is influenced by linguistic experience through psychological experiments. Their results showed that perceptual grouping was different depending on the native language of the participants. However, Jeon and Arvaniti (2016) used stimuli comparable to those of Iversen et al. (2008) and found neither a cross-linguistic difference of grouping between English, Greek, and Korean participants, nor a strong preference for trochees or iambs in any of the groups. This result is different from Iversen et al. (2008), who found a clear preference between different native language groups, making the findings from grouping studies less certain.

The studies on perceptual grouping suggest that native language affects rhythm perception, though not all of the studies show a perceptual preference for iambic or trochaic patterns. However, these studies have not yet shown whether native language influences the accuracy with which rhythmic structure is perceived, regardless of whether the rhythm is linguistic, musical, or tonal. This currently unanswered research question is examined in this study through the performance of perceptual experiments with sound stimuli. This study is inspired by Hannon and Trehub (2005), who demonstrated that rhythm perception can become more accurate with musical experience. In the study, they had USA participants, unfamiliar with Balkan music, listen to a 12-minute audio CD of computer-synthesised musical pieces of Balkan folk music with irregular 4+3/4 metre (non-binary rhythm), for two weeks. A second group of participants did not listen to the CD. The two experimental groups were then asked to rate the rhythmic difference between two sound files. As a result of the experiments, Hannon and Trehub (2005) showed that the experience of ethnic music allows participants to grasp the rhythm used. Considering their result that auditory experience boosts the ability to detect rhythm, it seems that linguistic experience can also make the rhythm perception sensitive; thus, the musical and linguistic experiences are examined and compared in this study (for more details on the study, see Section 1.5.)

The study of rhythm has been an interdisciplinary field, concerning linguistics, psychology, and musicology, which means that rhythm cannot be fully understood without insights from all three fields. This is what is undertaken in this study.

1.2 Early studies on native language and sound perception

1.2.1 Prior studies on phoneme and linguistic experience

Here, I discuss studies on L2 perception in more detail, because they demonstrate that our auditory perception can be influenced by the native language. L2 studies clarify the research questions of the current study on how our auditory perception, possibly including rhythm perception, can be influenced by linguistic experience. There are vast numbers of studies on this topic, too many to review here, so only a small selection of studies are discussed to showcase the scope of relevant research. In addition to the study by Goto (1971) mentioned in Section 1.1, Polka (1991) found that English speakers perceive Hindi dental stops and retroflex stops [ʈ] [ɖ] consonants [t̪] [d̪] as dental consonants. Furthermore, through an AX discrimination task, Werker and Logan (1985) found that English speakers have difficulty in distinguishing the voiceless dental stop /t̪/ and voiceless retroflex stop /ʈ/, especially when the temporal distance between the two stimuli in each AX trial increases. The study suggests that distance between a pair of sound stimuli has to be considered in experimental design, particularly because the task for participants in the current study is to distinguish rhythmic difference within a pair of sound stimuli. Comparable studies on vowels can be found in Beddor and Strange (1982), Bohn and Flege (1990), Stevens, Liberman, Studdert-Kennedy, and Ohman (1969). Perceptual difficulties with consonants are also demonstrated in Lisker and Abramson (1967): they show that English, Spanish, and Thai speakers perceive differences between /b/ and /p/ differently, due to differences in voice onset time (VOT) – the temporal interval between the release of a stop consonant and the onset of voicing. The result of the study is also supported by Abramson and Lisker, (1970), who showed that English speakers have difficulty distinguishing Spanish voiced consonants. Werker and Lalonde (1988) also support that voiced and voiceless consonants are perceived differently: Hindi speakers can distinguish small differences in VOT that English speakers had difficulty distinguishing. Werker and Tees (1984) showed that Salish speakers are more sensitive to the difference between glottalised velar and glottalised uvular sounds, /k'/ and /q', than English speakers. The studies mentioned in this paragraph show that sounds we are unfamiliar with from our native language are more difficult to perceive.

Werker and Tees (1984) also showed that English infants could distinguish the difference between glottalised velar and glottalised uvular sounds better than English-speaking adults, concluding that the influence of the native language on consonantal perception was found more clearly in adult groups than infant groups. Werker and Lalonde (1988) support this conclusion, demonstrating that Hindi speakers and English-speaking infants performed better in distinguishing the voiced bilabial stop /ba/, the voiced dental stop /d̪a/, and the voiced retroflex stop /ɖa/ than English speaking adults. Polka and Werker (1994) also showed that English-speaking, younger infants perceived German vowels more accurately than English-speaking, older infants. Eimas (1975), Lasky, Syrdal-Lasky, and Klein (1975), and Werker and Lalonde (1988) found that infants initially discriminate foreign phonemes that adults cannot discriminate. However, Werker and Tees (1984) investigated when English-speaking infants began to have difficulty distinguishing Spanish phonemes not used in their native language and found that their discrimination declined after 12 months of age. Through these studies with infants, the influence of the native language is clear: young infants can discriminate sounds that older infants and adults cannot, which demonstrates the influence of native language on the perception system. In the current study, I look at whether this influence extends to the perception of rhythm.

1.2.2 Prior studies on prosody and linguistic experience

A number of studies during the past two decades have shown the influence of native language on prosodic perception: these include, Dupoux, Pallier, Sebastian, and Mejler (1997); Dupoux, Peperkamp, and Sebastián-Gallés (2001); Dupoux, Sebastián-Gallés, Navarrete, and Peperkamp (2008); Lin, Wang, Idsardi, and Xu (2014); Lukyanchenko, Idsardi, and Jiang (2011); and Tremblay (2008).

Dupoux et al. (1997) found that French native speakers have difficulty discriminating words that differ in the position of stress, while Spanish speakers do not. According to Dupoux et al. (1997), this difficulty in detecting stress (known as *stress deafness*, after Peperkamp and Dupoux [2002]) is due to the fact that stress is not used in French to distinguish the meaning of words, while stress does have this function in Spanish. Stress is contrastive in Spanish and can change the meaning of words, for example: ˈsabana ('sheet') and sabana ('savannah'); and ˈlimite ('boundary'), limite ('[that] he/she limit'), and limiˈte ('limited'). Stress (or, more generally, alternations in prominence) are the basis of rhythm, as is discussed in more detail in Section 1.3. Thus, studies on stress deafness imply that rhythmic perception can be influenced by stress perception and, by extension, our native language. Stress and accent are discussed in greater depth in the following sections, because of the role they play in shaping rhythm perception in language.

1.3 Rhythm in language

As summarised by Nespor, Shukla, and Mehler (2011), linguistic rhythm can be defined in two ways:

(1) A regular alternative pattern of prominent and non-prominent syllables, or

(2) A regular isochronous pattern at a different level (stress, syllable, or mora), depending on the language.

Regarding the first categorisation, linguistic rhythm can be defined as the alternation of prominent and non-prominent element (where elements are a prosodic unit, such as syllables) groupings, such as feet, which consist of both prominent and non-prominent elements.

With respect to the second rhythmic categorisation, isochrony is an idea that language rhythms differ according to which units of speech have similar durations. For instance, according to the isochrony theory, inter-stress duration is considered to be constant in English, while inter-syllable duration is constant in French. Since the inter-stress interval in English is a unit that must exhibit isochrony, English is classified as a stress-timed language.

The concept of isochrony in speech has given rise to the idea of rhythm classes, stress-timing, syllable-timing, and mora-timing. However, both are questionable, having been refuted by studies in production and perception, as shown by Arvaniti (2009; 2012), among others. Isochrony and rhythm classes are discussed here because they are still widely adopted.

Pike (1945) mentions that Spanish rhythm consists of successive syllables and that English rhythm consists of stress-based feet. Abercrombie (1967) suggests the notion of isochrony for each rhythm category, i.e. the inter-stress duration is constant in stress-timed languages, while syllabic duration is constant in syllable-timed languages. Abercrombie (1967, p. 98) also explains as follows:

(i) ‘there is considerable variation in syllable length in a language spoken with stress-timed rhythm whereas in a language spoken with a syllable-timed rhythm the syllables tend to be equal in length’.

(ii) ‘in stress-timed languages, stress pulses are evenly spaced’.

English and Russian have been classified as stress-timed languages by Abercombie (1967), which implies that their rhythms are similar. Japanese, on the other hand, is classified as a mora-timed language whose moraic duration is constant, as discussed in Bloch (1950) and Jinbo (1980). Here, the word mora (μ) suggests a phonological unit that corresponds to a letter in Japanese (for a detailed explanation of mora, see Section 2.2.2).

Contrary to popular belief, many studies on timing failed to prove the isochrony, or constant duration of each unit (interstress interval or foot, syllable, or mora), as mentioned in Bertinetto (1989) and Arvaniti (2009), among others. For example, Bolinger (1965), Lea (1974), O'Connor (1965), and Shen and Peterson (1962) showed that inter-stress duration in English increases in proportion to the number of syllables, which means that inter-stress duration is not constant. Similarly, Borzone de Manrique and Signorini (1983) showed that syllable duration in Spanish, which is supposed to be a syllable-timed language, is not constant. Wenk and Wiolland (1982) demonstrated that the syllabic duration in French is not regular either, and Warner and Arai (2001) showed that mora duration is also not constant.

Despite the difficulty in finding evidence in favour of rhythm classes, many attempts, such as metrics and formulas, which show the durational variability of consonantal and vocalic intervals, have been proposed and tested using linguistic materials from a number of languages. Proposed metrics by Ramus et al. (1999) are ΔV (standard deviation of inter-vowel duration), ΔC (standard deviation of inter-consonant duration), and %V (percentage of vocalic intervals in the utterance). Other metrics that were developed to achieve precise rhythmic classification are VarcoV, by Dellwo (2006) (standard deviation of inter-vowel duration divided by the mean, multiplied by 100); VarcoC, by Dellwo (2006) (standard deviation of inter-consonant duration divided by the mean, multiplied by 100); nPVI-V, by Low, Grabe and Nolan (2000) (mean of the differences between successive inter-vowel duration divided by their sum, multiplied by 100); and rPVI-C, by Low, Grabe and Nolan (2000) (mean of the differences between successive consonantal intervals).

Arvaniti (2009) reviewed these metrics and argues that they are problematic, mentioning that one problem with metrics is that none of the studies could correctly classify all (or even the majority of) the languages tested in studies such as Grabe and Low (2002) . Similarly, in a cross-linguistic study by Arvaniti (2012), the %V score for German – a language said to be stress-timed – was lower than that of Italian, which is considered to be syllable-timed. Looking at the results in Arvaniti (2012), none of the metrics could classify multiple languages (English, German, Greek, Italian, Korean, and Spanish) into supposed rhythm categories (stress-timing or syllable-timing). Some of the metrics (%V, ΔC, nPVI, rPVi, VarcoC, and VarcoV) tested in Arvaniti (2012) could classify a language into a rhythm category, but these metrics failed to classify some other languages. Arvaniti (2012) found that substantial inter-speaker differences caused inconsistency in the results. Moreover, she found that metric scores can be affected by the type of data analysed (e.g. scores depend on whether spontaneous or read speech is analysed), and by the complexity of the syllables that happen to feature in a given speech sample. In conclusion, the study casts doubt on the use of metrics to classify rhythm classes and, by extension, to rhythm classes themselves, as no other evidence for rhythm classes is forthcoming.

Some evidence in favour of rhythm classes appears to originate from studies of language acquisition: Nazzi, Bertoncini, and Mehler (1998), Nazzi, Jusczyk, and Johnson (2000), and Nazzi and Ramus (2003) show that infants can correctly discriminate between stimuli from syllable-timed and stress-timed languages, such as English and French. In these studies, stimuli were manipulated by replacing each consonant and vowel with /s/ and /a/, respectively. The stimuli were manipulated in this manner, so infants could react only to durational differences in the stimuli. While experimental studies on isochrony failed to prove that linguistic rhythms are isochronous and can be categorised in rhythm classes, the studies on infant rhythmic perception show that stress-timing or syllable-timing can be discriminated from each other. However, Nazzi, Jusczyk, and Johnson (2000) also found that infants discriminate rhythm between varieties of the same languages, such as British English and American English, which undoubtedly belong to the same rhythm class. This result contradicts the idea that successful discrimination in these experiments is related to rhythm class. Arvaniti and Rodriquez (2013) suggest that the reason for successful discrimination (or lack thereof) may be the speaking rate differences between languages.

Considering this, it seems that the rhythm class and metrics are not reliable in defining rhythm in the current study. Facing problems involved in metrics and rhythm classes, Arvaniti (2009) proposes a way of thinking of linguistic rhythm based on a psychological perspective adopted by Fraisse (1963; 1982) and Woodrow (1951): namely, to consider rhythm as a binary pattern of prominence, in other words, not durational consideration but rhythmic pattern. Therefore, in this thesis, the definition of rhythm is an alternative pattern of prominent and non-prominent elements, a definition similar to the first option provided by Nespor, Shukla, and Mehler (2011).

1.4 The influence of native language on rhythmic perception

As previously mentioned in Section 1.1, there is evidence that rhythm perception is influenced by one's native language. This is illustrated most strongly in relation to the iambic-trochaic law (Hayes, 1995). When people listen to a series of sounds that alternate between long and short, listeners group them into iambs, i.e. they consider that the longer sound is group-final. In contrast, a series of alternating loud and soft sounds are preferentially grouped into trochees, i.e. the louder sound is considered group-initial. Many (though not all) studies show that these preferences are modulated by one's native language.

Also mentioned in Section 1.1, Iversen, Patel, and Ohgushi (2008) demonstrated that English and Japanese speakers perceive a rhythmic phrase differently. According to their study, in listening to the successive alternation of short and long sounds, English speakers tend to answer that the short sound is the beginning and the long sound is the ending of a group, while Japanese speakers preferred a long-short pattern than a short-long one. In considering their study, it appears that rhythmic perception can be influenced by language. Moreover, Bion, Benavides-Varela, and Nespor (2011) found that Italian adults considered high-pitched syllables as group-initial and long syllables as group-final, suggesting that fundamental frequency (high-pitched syllables) works as prominence similarly to that of amplitude.

However, some studies show that the rhythmic preference is not language specific: Hay and Diehl (2007) tested French and English speakers and found no differences in rhythmic grouping preference between the two groups. Additionally, they found that their participants' responses were comparable between speech and non-speech materials, suggesting that the same principles of rhythm perception apply to both linguistic and other modalities. In considering Hay and Diehl (2007), it seems that the iambic-trochaic law is a universal phenomenon not influenced by cultural aspect or modality. Furthermore, Jeon and Arvaniti (2016) showed that there were neither differences of grouping between English, Greek, and Korean speakers to perceive the sounds, nor a specific preference between them. In short, studies of the iambic-trochaic law suggest that it has an innate component, but one that can be affected by the native language.

1. Originally iambic-trochaic law focused on intensity and duration but, by testing both adults and infants, Bion et al. (2011) showed that pitch works similarly to intensity as a prominence; however, they also showed that infants did not prefer the long syllable at the end of a group and that they preferred the high-pitched syllable to be at the beginning of a group. Subsequently, the study suggests that it is a universal principle, as some aspects of it are found with infants, but it is also modulated by language, in that adult results differ from those of infants.

Furthermore, the role of the native language is shown in how speakers of different languages parse putative rhythm units, like syllables and mora. According to Otake, Hatano, Cutler, and Mehler (1993), French speakers perceive Japanese phrases as a succession of syllables, while Japanese speakers segment Japanese phrases based on mora (see Section 2.2.2 for a detailed explanation of mora). According to Cutler (2000), Japanese speakers perceive English /n/ as a mora, while English speakers do not, in other words, Japanese speakers perceive English syllables as moraic units. Subsequently, the English word *can* will be considered as two morae (ca + n) [kya+n] by Japanese speakers.

The studies discussed in this section did not examine whether native language improves rhythm perception; however, Hannon and Trehub (2005) demonstrated that musical experience of non-binary rhythm increases the accuracy of non-binary rhythm (irregular pattern of prominent and non-prominent units); in other words, rhythmic perception can be trained through experience. This notion was tested in this current thesis using the experimental design of Hannon and Trehub (2005), who examined the influence of musical experience on rhythm perception. As such, the experimental design of this thesis is based on their concept. Further details of their study are provided in Section 1.5.

1.5 Hannon and Trehub (2005): the role of experience on rhythm perception

This current study is based on the paradigm developed in Hannon and Trehub (2005), who examined the relationship between rhythmic perception and exposure to rhythm (rhythmic experience).

As outlined in Section 1.1, the aim of Hannon and Trehub's (2005) study was to examine whether exposure to the non-binary rhythm (4+3/4) used in Balkan dance music could increase accuracy, operationalised as the ability to detect deviations from this type of non-binary rhythm. To test this hypothesis, Hannon and Trehub (2005) used a similarity judgment task, in which participants were asked to judge whether there was a rhythmic (metric) difference between two music clips (see Figures 1.1 and

1.2 on the design of stimuli). Forty participants (25 women and 15 men), aged 18–35 years old and all raised in North America, took part. They heard two stimuli and were asked to rate the rhythmic differences between them. These rhythmic differences in the stimuli are illustrated in Figures 1.1 and 1.2. and the accompanying text.

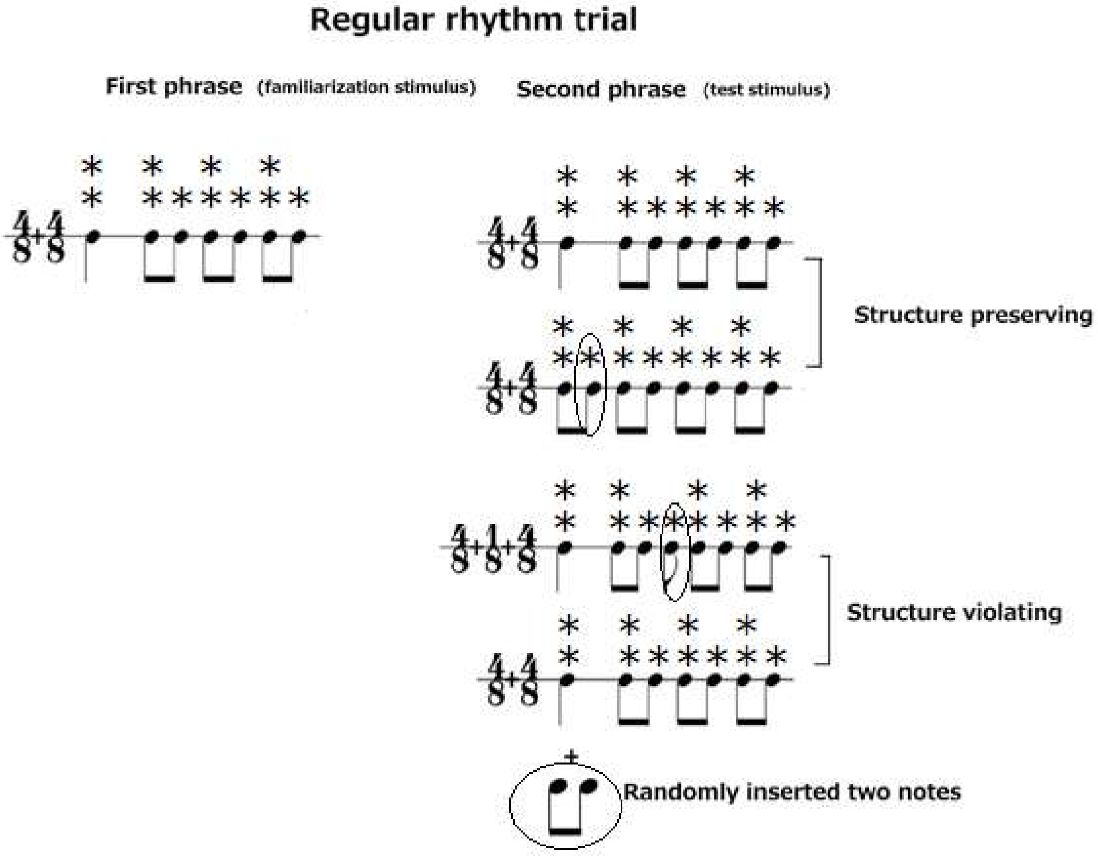


Figure 1.1: Experimental design of binary rhythm trial in which participants were asked to judge the rhythmic difference between the first phrase and the second phrase. Circled notes make the test stimuli different from the familiarisation stimulus. The stimuli are available at the following link: https://www.pnas.org/content/102/35/12639

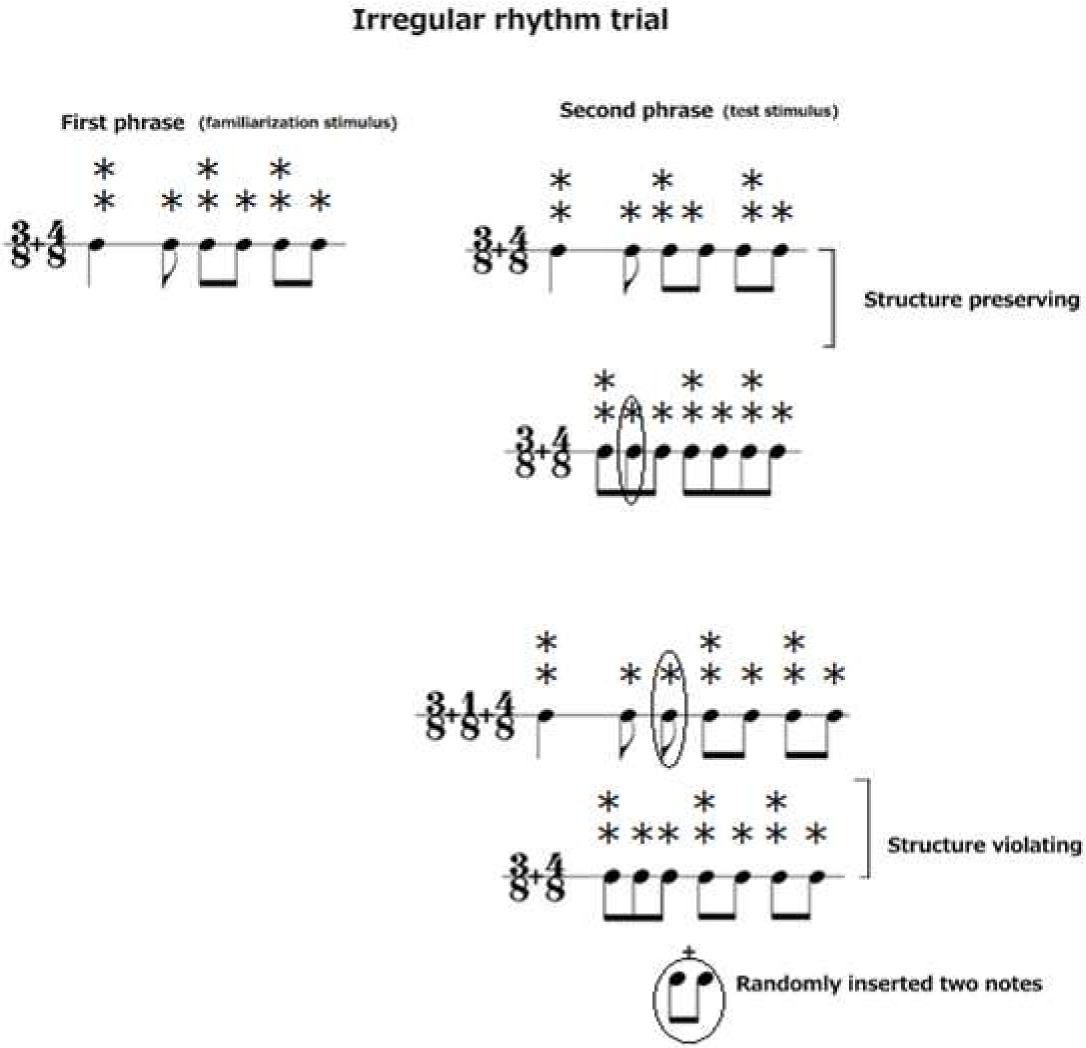


Figure 1.2: Experimental design of non-binary rhythm trial in which participants were asked to judge the rhythmic difference between the first phrase and the second phrase. Circled notes make the test stimuli different from the familiarisation stimulus. The stimuli are available at the following link: https://www.pnas.org/content/102/35/12639

There were two types of trial. The first involved binary rhythm, where participants rated the rhythmic difference between the first musical phrase (familiarisation stimulus), which consisted of a binary rhythm (4+4/8), with a second phrase (test stimulus), different from the familiarisation stimulus due to the insertion of eight notes (♪), as seen in Figure 1.1. The second set of trials involved a comparison between the familiarisation stimulus, whose rhythm was non-binary (4+3/4), with the test stimulus, whose rhythm differed from the familiarisation stimulus in that eight notes were inserted, as can be seen in Figure 1.2.

There were two types of test stimuli in their experiment: structure-preserving and structure-violating stimuli. The structure-preserving stimuli were, as the name suggests, similar to the familiarisation stimulus in the sense that the metre of the familiarisation stimulus and the structure-preserving stimuli were identical (see Figure 1.1. for an example). Contrary to this, structure-violating stimuli were different from the familiarisation stimulus due to two additional notes, which violated the rhythm structure of the familiarisation stimulus. The position of the inserted notes in the test stimulus was random (see Figures 1.1 and 1.2).

Participants were asked to rate the rhythmic difference between the two sound files on a Likert scale from 1 (very similar) to 6 (very different). Hannon and Trehub (2005) then calculated what they called *the rhythmic accuracy* of each participant, by subtracting the participant's average rating for the structure-preserving trials from their rating for the structure-violating trials. Participants took the test and were then divided into a control group and an experimental group. Those in the experimental group were asked to listen to a CD at home. Each CD contained five recordings of dance music with a non-binary rhythm. The total duration of recordings was approximately 10 minutes.

After one to two weeks of exposure to Balkan music with a non-binary rhythm, a second experimental session, identical to the first, took place. The accuracy with which participants in the experimental group detected deviations in stimuli with a non-binary rhythm was higher in the second session than in the first session (though it remained significantly lower than accuracy with binary rhythm). This result shows that exposure to the non-binary rhythm in Balkan music enhanced the participants' perception of such rhythms. It also indicates that experience with different types of rhythm affects rhythm perception. As such, the experimental design of Hannon and Trehub (2005) is suitable for examining the influence of native language rhythm and musical rhythms on rhythm perception.

1.6 Possibility of empirical influence on rhythm perception

In considering the findings in the literature reviewed in the previous sections, one hypothesis arising is that linguistic rhythm in native language influences rhythm perception, both in language and other modalities. The main subject of this thesis is to examine this hypothesis through experiments with Japanese, English, and Russian speakers. Moreover, as Hannon and Trehub (2005) also suggest, musical experience can affect rhythm perception; thus, musical experience is also examined.

One hypothesis examined through perception experiments is whether native language and/or musical experience affect sensitivity to binary and non-binary rhythms.

Following on from Hannon and Trehub (2005), here the binary rhythm is defined as a regular alternation of prominent and non-prominent elements similar to the musical metre 2/4; for example, $\overset{x}{x}x\overset{x}{x}x$ $\overset{x}{x}x\overset{x}{x}x$, where ‘$\overset{x}{x}$’ stands for a prominent element, higher or louder than ‘x’, which is a non-prominent element. For ease of exposition, this will be referred to as binary rhythm for the remainder of this thesis. Meter within a group of elements with a $\overset{x}{x}x\overset{x}{x}xx\overset{x}{x}xx\overset{x}{x}xx$ (2+3/4) pattern is referred to as non-binary rhythm for the remainder of this thesis.

Rhythmic perception is of great interest, as cultures differ in terms of the rhythms they use in traditional forms of music, and languages also differ in terms of how regular or irregular their rhythm is. English speakers are familiar with binary rhythm in music and language and non-binary rhythm is relatively common in Japanese music and language; however, although Russian rhythm is relatively binary, similar to English, rhythms in some traditional Russian music are non-binary, similar to that found in Japanese music (see Sections 2.1, 2.2, and 2.3 for further detail on each linguistic rhythm and sections 3.3, 3.4, and 3.5 for musical rhythm). Although it would be easy to compare English and Japanese speakers to observe differences in rhythm perception, the problem is that their musical backgrounds, which can influence rhythm perception, are entirely different. Thus, it is not sufficient to show the linguistic and musical influence on rhythm perception simply by comparing English and Japanese speakers. The third group – Russian speakers – supports the comparison of linguistic and musical influence, as their linguistic background is similar to English speakers, whose native language has binary rhythm. However, their musical background is similar to that of Japanese speakers, as both are familiar with non-binary musical rhythms used in folk and traditional music, and both cultures place emphasis on musical education for all. It is also important to consider that Russian speakers tolerate a lapse, which can differ to English rhythm (see Section 2.1.2).

The cross-cultural comparison of rhythm is complicated, due to the fact that both linguistic and musical experiences can affect rhythmic perception; however, if the Russian speaker's rhythmic perception is similar to the English speaker's, rather than the Japanese speaker's, it would suggest that linguistic experience affects rhythm perception more than musical experience, as English and Russian speech rhythms are more alike than when compared to Japanese. Equally, if the Russian speaker's rhythmic perception is closer to that of the Japanese speaker's, rather than that of the English speaker's, it would mean that musical experience affects rhythmic perception more than the rhythm of one's native language. Similarly, comparing musicians and non-musicians helps to assess the degree of influence of linguistic and musical experience. The details of linguistic rhythm and musical rhythm are discussed in greater depth in chapters two and three.

Chapter 2: English, Japanese and Russian rhythms

2.1 English

Understanding stress is necessary for this study as it is, according to some accounts, the basis of linguistic rhythm (Arvaniti, 2009; Dauer, 1983; Hayes, 1995). Some definitions of stress accent (henceforth referred to as 'stress' for short) are provided by Hyman: 'Stress = metrical structure present lexically (e.g. at word level)' (2001a, p. 256); 'The stress-bearing unit is the syllable...' (2009, p. 217); and:

> A language with stress is one in which there is an indication of word-level metrical structure meeting the following two central criteria: Every word has AT LEAST AND AT MOST one syllable marked for the highest degree of metrical prominence. (2011, p.)

As presented in the definitions above, stress is a property of word rather than phrasal level and is different to sentence accent (sentence accent here means the most prominent syllable within the whole sentence, being placed on the primary word stress of a content word). As rhythm is a pattern of stressed elements (stressed syllables in English) and non-stressed elements in general, it is necessary to consider the feature of English stress.

2.1.1 Stress accent in English

In an early study, Fry (1958) argued that fundamental frequency is the most important cue to detect the stressed syllable. He mentions that, in English, the position of stress can change the meaning of a word; for example, 'súbject' is a noun or adjective, but 'subjéct' is a verb. In this example, depending on the position of stress, the word 'subject' can be a noun or a verb (the diacritic marks, or acute accents, suggest stressed syllables). This is significant to the current study because, as is shown in Section 1.4, one is sensitive to linguistic elements that can influence the meaning of words, meaning that English speakers are expected to be sensitive to stress. Bolinger (1985) and Morton and Jassem (1965) concluded similarly to Fry (1958), pointing out that pitch is the most important cue to English stress accent.

Beckman (1986) presented two experiments to examine acoustic correlates and acoustic cue: one was a production experiment to study the acoustic correlates of stress in English and pitch accent in Japanese (see Section 2.2.1 for more detail on pitch accent) and the other was designed to examine perceptual cues to stress in English and pitch accent in Japanese. In the first experiment with production, she used bisyllabic words, which can be trochaic (stressed syllable first, followed by an unstressed syllable) or iambic (unstressed syllable first, followed by a stressed syllable), such as *permit, contract, digest, subject, object*. Native speakers of American English read each word in two different ways (trochaic and iambic) and the recorded sounds were further analysed to examine if first stress and second stress were produced similarly in terms of duration, amplitude, and fundamental frequency. The results showed that English stress was expressed by all correlates (i.e. pitch, loudness, and duration). In another experiment by Beckman (1986), the perceptual cue to English stress was examined: stimuli were similar to the first experiment in that bisyllabic words, which can be either trochaic or iambic, were used. Beckman (1986) recorded bisyllabic words then manipulated them so that the first or second stressed syllable was a sole correlate (i.e. pitch, duration, vowel quality, or loudness). Participants were asked to judge if the stress in the stimulus was a first syllable or second syllable. The results showed that pitch, duration, and vowel quality are equally important cues to detect stress, while loudness is a less robust, but still reliable, cue for English speakers. Through comparative experiments between English and Japanese speakers, Beckman concluded that all cues are important in English, while pitch is the sole robust cue in Japanese.

The early studies by Bolinger (1985), Fry (1958), and Morton and Jassem (1965) mentioned previously, did not consider the difference between sentence stress, whose prominent cue is pitch, and stress accent, which is a property of a word. Subsequently, recent studies that discriminate stress from sentence accent supersede these earlier works. The misunderstanding of pitch as a cue for stress is illustrated here:

> In one-word utterances, however, pitch excursions are more likely to be interpreted in terms of the sequence at a nuclear accent, as in experiment of Fry (1958) showing the salience of the F0 contour in cueing stress in pairs as pérmit versus permít. This is probably the major source of the common misunderstanding in the experimental literature that F0 excursion is a direct acoustic correlate of the feature "stress," a misunderstanding that has been incorporated into several standard textbooks… (Beckman & Edwards, 1994, p. 13)

To cope with the difficulty caused by sentence accent, Sluijter and van Heuven (1996) compared focused words (words with sentence accent) with non-focused words (a words without sentence accent) in an American English corpus elicited from six speakers. They analysed the fundamental frequency, duration, and intensity of stressed and unstressed syllables in each group (focused and non-focused words). The result demonstrated that fundamental frequency, duration, and intensity were similarly important acoustic correlates in focused words, while fundamental frequency was the poorest correlate in non-focused words. The most effective correlate in non-focused words was duration, and the second most effective correlate was vowel quality, while intensity was the second poorest correlate in the non-focused words category.

Kochanski, Grabe, Coleman, and Rosner (2005) examined how English speakers mark the stressed syllable in sound data from the IViE corpus, which consists of three styles of speech (sentences, read story, and retold story) and a range of dialects: Belfast, Bradford (speakers of Punjabi heritage), Cambridge, Dublin, Leeds, London (speakers of Jamaican heritage), and Newcastle. Stressed syllables tended to be rated as loud or long, but listeners did not notice or mark that the stressed syllable was accompanied by a pitch shift, even though stressed syllables were higher than non-stressed syllables. Kochanski et al. (2005) explained that the neglect of pitch by listeners is due to many other non-stressed syllables also being high.

Considering the recent studies on cues, duration, and vowel quality play important roles in English stress, while loudness also works as both correlate and cue. However, contrary to earlier views, the pitch is not a direct correlate of English stress.

2.1.2 English rhythm

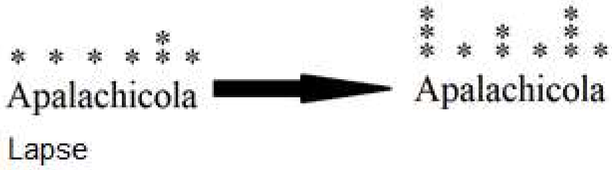
In this thesis, the definition of rhythm is a pattern of accented and unaccented elements (stress accented syllables in English and Russian, pitch accented mora in Japanese, and accented notes in music). Therefore, in English, the linguistic rhythm consists of accented and unaccented syllables.

English prefers a binary alternation of a stressed and an unstressed syllable, as suggested by Giegerich (1985), Gussenhoven (1991), Hayes (1984), Hogg and McCully (1987), Kiparsky (1979), Liberman and Prince (1977), Nespor and Vogel (1989), and Selkirk (1984). This means that clash, an irregular rhythm caused by a successive accented syllable, is avoided in English, as demonstrated in example (2.1). In the current study, the term ‘irregular rhythm’ refers to the clash and lapse. Hayes (1984), Prince (1983), and Selkirk (1984) all mention eurhythmy, which triggers the stress shift demonstrated in example (2.1); the ideal linguistic rhythm (eurhythmy) is an alternation of stressed and unstressed syllables in English, and accent position may be moved to achieve this.

(2.1)

Clash is a succession of the stressed syllables causing a rhythmic irregularity. This unfavourable rhythm is avoided by a stress shift like that seen in example (2.1) from Liberman and Prince (1977): the clash on the left-hand side of the example above is avoided by changing the position of an accented syllable. As the stress shift is optional, it does not necessarily occur every time there is a clash, for instance, an experiment by Shattuck-Hufnagel, Ostendorf and Ross (1994) showed that only 68% of clashes were avoided by a stress shift. Selkirk (1984) claimed that syllable lengthening and pausing may occur to avoid a clash, but Cooper and Eady (1986) did not find evidence of this in their studies. Grabe and Warren (1995) examined whether the stress shift phonetically occurs to avoid the stress clash or if listeners are biased to hear a shift: they asked English participants to identify stressed syllables in stimuli and found that listeners reported hearing a stress shift, even when they heard a phrase that acoustically contained a stress clash. This suggests a possibility that British English speakers may consider non-binary rhythm as binary rhythm in the current study.

(2.2)



Another type of irregular rhythm, the lapse, is a succession of weak, unstressed syllables – this is the irregular rhythm that violates eurhythmy in English. The alternation of prominent and non-prominent syllables results in the occurrence of a beat addition to avoid the lapse, as shown by Selkirk (1984) (see example 2.2). According to Shattuck-Hufnagel and Turk (1996), most words in English, with the exception of prepositions, articles, pronouns, and conjunctions, have at least one stressed syllable and thus long lapses are relatively rare in English.

Wasow, Levy, Melnick, Zhu and Juzek (2015) demonstrated that clash and lapse are avoided in English, not only by changing the position of an accented syllable or adding an additional accent, but by omitting or adding an optional ‘to’. This is demonstrated in the example below from Wasow et al. (2015), with stressed syllables marked in bold:

(2.3) And **one** of the **best ways** to **do** it **is** (to) **break bread** with **them**.

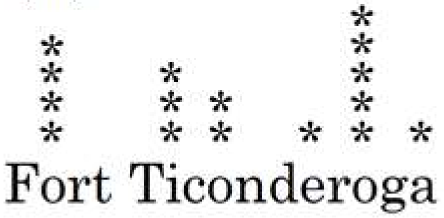
(2.4) **All I** can **do is** (to) con**ti**nue to be**have** in a **way** that **earns** your **trust**.

In (2.3), omitting ‘to’ would cause a clash, while including ‘to’ in (2.4) causes lapse. Wasow et al. (2015) found whether the optional ‘to’ occurs depends on if the omission or addition of it causes clash or lapse: English speakers in American English omit or include ‘to’ to avoid clash and lapse, for instance, in (2.3), English speakers tend to include ‘to’ to avoid clash while they omit ‘to’ in (2.4) to avoid lapse.

Another rhythmic feature of English discussed in Hayes (1984). According to Hayes (1984), the rhythmic structure in (2.5) is superior to that in (2.6), because it implies a 4/4 rhythm used in English music and is the ideal rhythm in the English language (see Section 3.2 for further explanation on rhythm in English music).

(2.5)

(2.6)



According to Cutler and Carter (1987), in English, the first syllable of a disyllabic content word tends to be an accented syllable, which is then followed by unaccented syllable. It is shown by Morton and Jassem (1965) and van Heuven and Menert (1996) that English speakers perceive stress on the first syllable, even when intensity, pitch, and duration of syllables in the word are identical.

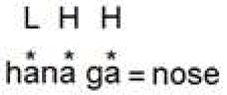
The discussion in this section demonstrates that the English rhythm is binary, and that the binary rhythm is maintained, not only by stress shift, but also by the biased perception of English speakers.

2.2 Japanese

2.2.1 Pitch accent in Japanese

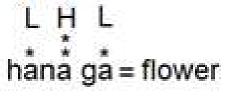
When looking at the basics of Japanese pitch accent, the lexical meaning of words is different depending on the position of the accent. For example, the Japanese words ‘hana ga’ can have different meanings depending on the position and presence of accent: ‘hana’ means flower or nose, and ‘ga’ is a particle that suggests that the preceding word is a subject in a phrase, as shown in Examples (2.7) and (2.8).

(2.7)



L H H
* * *
hana ga = nose

(2.8)



L H L
*
* * *
hana ga = flower

The pitch of ‘hana’ is identical in both meanings (flower and nose); however, the pitch of ‘ga’ following ‘hana’ has to be low to express ‘flower’. The ‘na’ preceding the low ‘ga’ is perceived as accented mora, while the pitch of ‘ga’ following ‘hana’ has to be high similarly to the preceding mora ‘na’.

There was controversy about the acoustic cue of Japanese pitch accent: a study by Neustupný (1966) suggested that pitch may not be a primary cue to detect an accent in Japanese, as the accented high vowels /i/ and /u/ are devoiced when they are between voiceless consonants, and a devoiced vowel does not have f0 information. Neustupný (1966) found that Japanese speakers perceive some of the voiceless vowels as accented through examining pitch contours and intensity of recorded Japanese phrases. In a spoken phrase like ‘fue o fuku’ (play a recorder), /fu/ is perceived as accented mora, although this /fu/ is usually devoiced (see example (2.9)). According to Neustupný (1966), the fundamental frequency of the devoiced /fu/ is impossible to perceive, as the mora is devoiced, but Japanese speakers perceive it to be prominent.

(2.9)

fue o fuku. (playing a recorder)

This phenomenon of a voiceless mora perceived as an accented mora is called ‘late fall’ or ‘ososagari’. Considering this phenomenon, in which Japanese speakers find accent on a voiceless vowel, Neustupný (1966) hypothesises that intensity, rather than pitch, can be an important cue to finding an accent in Japanese, because the fundamental frequency of a voiceless vowel is not perceivable.

Sugito (1972) showed that quick and big f0 fall at the beginning of mora that are proceeded by an unaccented (if voiceless) vowel and is a reason for the late fall. Even if the accented mora is devoiced and does not have f0, if the f0 of the following mora quickly falls, the voiceless mora is perceived as accented. Hasegawa and Hata (1992) supported this idea by Sugito (1972) by replicating the study, and both Beckman (1986) and Weitzman (1970) also mention that f0 is a primary cue to detect Japanese accent. Bybee et al. (1998, p. 277) describe Japanese pitch accent as: ‘A pitch-accent system is one in which pitch is the primary correlate of prominence…’ Furthermore, Sugiyama (2011) states that, in Japanese, loudness is a secondary cue and that f0 is the primary.

In short, the major differences between Japanese and English accents are that duration and vowel quality, while cues in English, cannot be cues to perceive the Japanese accent, and that f0 is the primary cue in Japanese, but not in English.

As suggested by Hyman (2006), in Japanese, morae of content word and function word can all be unstressed. In contrast, Shattuck-Hufnagel and Turk (1996) mention that content words in English and Russian have at least one accent. Hayashi (1982) showed that approximately half of Japanese words are non-accented and Tanaka and Kubozono (1999) showed the rate of Japanese nouns with an accent and the position of accented mora, as shown in Table 2.1. It is also clear from Table 2.1 that many Japanese words do not have an accent. Percentages in the table are approximate.

Table 2.1: Rate of accented and unaccented mora and the position of accented mora, from Tanaka and Kubozono (1999), with example words.

Moraic number of a word	No accent	First mora	Second mora	Third mora	Fourth mora
1	30% Example: 差 [sa] (difference)	70% Example: 矢 [ˈja] (arrow)			
2	15% Example: 鹿 [sika] (deer)	65% Example: 価値 [ˈkati] (value)	20% Example: 時 [toˈki] (time)		
3	50% Example: 孤独 [kodoku] (loneliness)	40% Example: トンビ [ˈtonbi] (black kite)	10% Example: 鍛冶師 [kaˈzis] (blacksmith)	5% Example: 痺れ [sibiˈre] (paralyze)	
4	70% Example: 新人 [sinzin] (beginner)	10% Example: 真実 [ˈsinzi͡ts] (truth)	10%< Example: 麒麟児 [kiˈrinzi] (genius)	10% Example: 淋しい [samiˈsiː] (lonely)	5% Example: タケノコ [takenoˈko] (bamboo shoot)

2.2.2 Japanese rhythm

It is widely believed that Japanese rhythm follows mora-timing. When first considering timing, Japanese is often regarded as a mora-timed language, although there are also many studies that deny this. Initially, Bloch (1950) and Jinbo (1980) both mentioned that Japanese is a mora-timed language, meaning the inter-mora duration tends to be regular. However, there are controversies over whether Japanese is mora-timed, similar to controversies over stress-timed languages previously discussed in Section 1.2.

Jinbo (1980) describes how *onsetsu* (mora) is produced in constant duration, regardless of whether a mora consists of consonant (C) + vowel (V), or solely of V. He goes on to mention that a phrase consisting of twice as many morae than another phrase will be perceived as twice as long, and also emphasised that Japanese listeners perceive each mora as regular. Ladefoged (2001) also mentions that mora production is constant in terms of duration. Experimental studies examined the regularity of mora, such as that of Han (1962), which showed that moraic duration is regular. Hattori (1980) argues that the second half of a long vowel (mora nasal) and a geminate obstruent are counted as a single mora, respectively.

The moraic count is different from syllabic count and can be summarised as follows:

(i) CV = one mora where V is a short vowel or diphthong

(ii) CV: = two moras, where V: is a long vowel

(iii) CVC = two moras, where the second C is either a nasal or the first half of a geminate

In other words, one mora is CV, V (single vowel or second half of long vowel), N (mora nasal), G (first half of geminate). The second part of a long vowel /:/, mora nasal /N/, and a geminate obstruent /G/ cannot be a syllable, but, in the Japanese moraic system, they are counted as a single mora, respectively. These three morae make mora-counting different from syllable-counting, and thus the term ‘non-syllabic mora' will be used in this study to describe the second half of a long vowel /:/, mora nasal /N/, and a geminate obstruent /G/). In the current study, the word 'syllabic' means CV or V morae and 'non-syllabic' refers to the second half of heavy syllables, i.e. the second half of long vowels, coda nasals, and the first half of geminates.

Otake (1988; 1989a) and Port et al. (1980) both indicated that adjacent segments would affect mora duration and that vowel duration depends on the preceding or following consonant. Port et al. (1980) conducted two experiments with Japanese and Arabic speakers to examine whether vowels following geminate consonants become shorter to compensate and keep moraic duration constant. Their results showed that, in Japanese, vowel length is compensated, and that the duration of two-syllable test words was constant throughout the experiments; however, this compensation was not found in Arabic. Minagawa and Kawai (1999) write that the duration of Japanese mora is relatively constant regardless of the following and preceding vowel and consonant. Sato (1993) examined the length of the vowel before a voiceless or voiced stop in English, Korean, and Japanese; it was demonstrated that, particularly in Japanese, the vowel length tends to be constant regardless of the influence of a voiced or voiceless stop. Port et al. (1987) investigated the hypothesis that a mora in a word will be shorter if another mora in the word is longer to balance the duration within the word; they showed a positive linear relationship between the total duration of a word and number of mora.

However, this does not necessarily suggest that Japanese moraic duration is constant, as pointed out by Warner and Arai (2001): even if Japanese speakers adjust the duration of consonants and vowels to balance inter-mora duration, the structure of mora (i.e. CV or V structure) makes them relatively constant when compared to syllables, which can have more complicated structures, such as CVC or CCV, as found in Germanic or Slavic languages. Moreover, Sato (1995) suggests that Japanese pitch accent – which is not expressed by lengthening, and is different from English and Russian, whose accented syllables are longer than unaccented syllables – does not make a durational difference between unaccented and accented mora. Fujisaki, Hirose and Sugito (1986) also showed that the durational difference between Japanese accented mora and unaccented mora is small, but it does not necessarily mean that the Japanese moraic duration is regular. However, these studies did not compare accented and non-accented morae statistically.

Although, the studies mentioned above support the view that Japanese moraic duration is constant, other studies found no such evidence. Arai (1999), Cambell and Sagisaki (1991), Han (1962) , and Sato (1993) all showed that long vowel /V:/ is not necessarily twice as long as other morae, which is as expected, given that it is counted as two morae in Japanese. Additionally, Arai (1999), Cambell and Sagisaki (1991), Han (1962), and Sato (1993) all demonstrated that durations of non-syllabic morae are different from the durations of other morae. Arai and Greenberg (1997) deny the possibility of mora-timing through showing that the variability of moraic duration in Japanese is too high, concluding that the perception of mora-timing is an illusion.

There are not many experimental studies covering the perception of mora-timing. Kato (1999) and Kato et al. (1997) investigated the durational perception of mora. These studies examined whether durational changes of vowels or consonants are acceptable to Japanese listeners and both experiments asked participants whether moraic duration in a stimulus was constant. They discovered that Japanese speakers do not perceive the duration of mora as constant when they listen to irregular sound patterns of mora. The results suggested that Japanese listeners perceive moraic duration to be irregular when the sounds are slightly shortened or lengthened; this contradicts early studies, such as Jinbo (1980), which suggested that Japanese speakers perceive a mora as a regular sound unit, even when the duration of a mora is not perfectly constant. In other words, the reason why Japanese has been labelled a mora-timed language is not because Japanese speakers perceive moraic duration to be constant.

The problem with the notion of rhythm classes, defining linguistic rhythm as consisting of a regular durational pattern, is that studies failed to prove whether languages are categorised into rhythm classes in production or in perception, as demonstrated by Arvaniti (2009). As a result, Section 1.2 defined rhythm as a regular pattern of prominent and non-prominent linguistic elements.

One difficulty is that Japanese rhythm cannot be categorised in the same way, as there is no stress to create a rhythmic structure. Although Japanese uses pitch accent instead, it is not certain whether pitch accent is prominent in a similar way to that of stress accent, which is considered to be the basis for creating rhythm. Subsequently, in the current study, it is necessary to consider whether Japanese pitch accent can be considered prominent in creating rhythm.

Another problem in defining Japanese rhythm is that, even if pitch accent does have rhythmic prominence similar to that of stress accent, many Japanese words do not include pitch accent, as shown by Tanaka and Kubozono (1999). This means that it is systematically difficult for Japanese to maintain a binary pattern of alternation of prominent and non-prominent morae, due to the high frequency of unaccented morae to keep the rhythm. Before tackling Japanese rhythm in more detail, we will need to examine the basics of Japanese rhythm.

Furthermore, it is not clear whether Japanese pitch accent can be considered the basis of rhythm in a similar way to English stress accent. In Japanese, instead of the stress accent used in English, pitch accent can change the meaning of words. Depending on pitch, the meaning of a word is different from another word consisting of identical phonemes, e.g. kámi (God) and kamí (paper) (see Section 2.1.1 for more detail on stress accent in English). Pitch-accented mora might be considered prominent; however, pitch-accented mora may also be louder than unaccented mora, and thus Japanese pitch accent can be perceived as a strong sound. This is supported by Fujisaki (1986), who compared the amplitude of Japanese and English accents and demonstrated that intensity of a pitch-accented mora is greater than that of unaccented mora (though the amplitude differences between unaccented and accented syllables in English was larger). If Fujisaki's conclusion is correct, pitch accent may be a plausible basis for rhythm in Japanese. Also, Fujisaki (1986) suggests that.

It is worth considering whether pitch accent can work similarly to English stress accent to form a linguistic rhythm, since the phenomenon of mora-timing is dubious, as previously discussed; however, Japanese correlates are different from correlates for the English stress accent. Considering the previously mentioned findings of Fujisaki (1986), the Japanese accent seems similar to the English accent, in that it is louder and accompanied by a pitch shift. Beckman (1986) compared English and Japanese correlates through her experiment and confirmed that correlates for Japanese accent is different from English accent in the sense that amplitude and duration of accented mora were not significantly different from those of unaccented mora; however, fundamental frequency, amplitude, and duration of stressed syllables in English were clearly different from those of unstressed syllables (see p. 20 for more detail on Beckman's study). The study by Fujisaki (1986), adopting an identical method to Beckman (1986), demonstrated that pitch is a correlate in both Japanese and English, but loudness and duration are significant correlates for English stress accent, not Japanese. However, Haraguchi (1991) and Tajima (1998) – both Japanese-English bilingual linguists – argue that the English stress accent and Japanese pitch accent both play roles of prominence, which can be used as the basis of linguistic rhythm; they mention that both English stress accent and Japanese pitch accent are prominent in a similar manner within a phrase.

Given that Japanese rhythm might be a combination of accented and unaccented morae – a pattern of prominent and non-prominent morae – the lapse is frequently found in Japanese. For example, there are only two accents in the Japanese long phrase (2.10).

(2.10)

efferu tou no ué karano késhiki

[eɸ.ɸerɯtounouˈekaranoˈkesiki]

'View from the top of the Eiffel tower.'

The reason for the frequent lapse in Japanese is that many Japanese words do not have an accent, whereas English and Russian content words contain at least one stress. The feature of the Japanese pitch accent is that it is not obligatory for each Japanese word. This is also the feature of pitch accent language in which the content words do not necessarily have an accent, as mentioned by Hyman (2006).

It is possible to find long lapses in Japanese, particularly in compounds and in phrases that contain a specialty, something that is referred to as specialist accent. According to Inoue (1998), when a specialist talks about their specialty, words concerning the specialty can lose the accent in Japanese. Example (2.11) of a rider talking about a bike demonstrates what happens when a specialist omits the pitch accent in Japanese.

(2.11)

‘báiku’ [ˈbaiku] (bike) → baiku [baikɯ]

Clash is systematically impossible in Japanese because pitch falling to the bottom of a register is a cue to detect a Japanese accent, as mentioned in the previous section. This means that pitch cannot fall further without a pitch reset.

Meanwhile, movement, or the deletion or addition of an accent or a word is preferred to make a linguistic rhythm binary in English (as mentioned in Section 1.2); such arrangements are not used to make rhythm binary in Japanese. One example of the removal of accent is, ‘kissaten’ (coffee shop), a compound of ‘kissa’ (to drink tea) and ‘tén’ (shop).

(2.12) ‘kissaten’ [kis.satenɴ] (coffee shop)

In standard Japanese, 'kissaten' can be pronounced in three ways: 'kissáten', 'kissatén', or 'kissaten'. Some Japanese speakers prefer the first pronunciation, but the other two patterns are also found in daily conversation. Kubozono (1995; 1997) tried to identify a rule or condition in these variants, hypothesising that the position of the accent would be decided to avoid irregular rhythms, such as lapse. However, by comparing pronunciations of compounds by Japanese speakers, Kubozono (1995; 1997) revealed that Japanese speakers did not consider the position of accent, nor rhythm. Summarising this, one may consider that there is seemingly no repetitive pattern of prominent and non-prominent morae in Japanese, which might suggest that there is no rhythm either.

However, some other studies suggest that Japanese has a trochaic bimoraic foot structure. Kurisu (1994) found that Japanese speakers prefer to divide a word before odd-numbered morae than even-numbered morae, which would mean that Japanese speakers prefer a bimoraic foot structure. Sakano (1996) argues that the Japanese verse-like phrase tends to have a two-morae feet structure, as shown in example (2.13), although this idea is not based on experiments or relevant past studies. Kozawa (2000) also argues that the bimoraic foot is common in Japanese, but this study was not based on experiments or studies either.

Similarly, Poser (1990) shows that many word-formation patterns in Japanese are based on bimoraic foot, for instance, most onomatopoeia. One example of this is 'Kirakira' and 'pikapika', which both mean shiny, and the foot structure is the same as (2.13), where the symbol 'μ' suggests a mora.

(2.13) [μμ][μμ]

Shortened Japanese words also tend to have a bimoraic structure, as shown in (2.14).

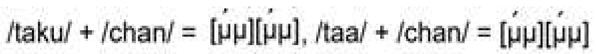
(2.14) tako-wasa < tako (octopus) + wasabi (wasabi).

gaku-wari < gakusei (student) + waribiki (discount)

In addition to this, Poser (1990) shows examples of trochaic bimoraic feet patterns in Japanese nicknames, although many examples are not trochaic bimoraic, but simply bimoraic due to a lack of consideration about prominence. The examples above show that the Japanese prefer bimoraic segmentation, bimoraic abbreviation, and onomatopoeia, but they do not consider whether the bimoraic structure is trochaic or iambic.

The example taku chan (2.15) is a possible nickname for Takumi, a male name whose rhythm is trochaic bimoraic feet. Poser (1990) states that this type of rhythm is frequently found in Japanese. Here, the trochee means that the initial mora of the bimoraic pair is pitch accented.

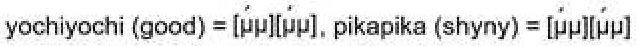
(2.15)



/taku/ + /chan/ = [μ́μ][μ́μ], /taa/ + /chan/ = [μ́μ][μ́μ]

Also, reduplications and onomatopoeia have a similar trochaic structure in Japanese. Example (2.16) demonstrates reduplication ‘yochiyochi’ (totteringly) and onomatopoeia ‘pikapika’ (glitter) with bimoraic trochaic structures.

(2.16)



yochiyochi (good) = [μ́μ][μ́μ], pikapika (shyny) = [μ́μ][μ́μ]

It is shown by Mazuka, Kondo and Hayashi (2008) that these types of trochaic reduplication and onomatopoeia are frequently found in infant-directed vocabulary, in Japanese. Similarly, Hayashi and Mazuka (2017) examined whether Japanese infants prefer these trochaic rhythms, through a series of experiments with Japanese infants between eight and 10 months old. They discovered that Japanese infants reacted to trochaic rhythm used in Japanese phrase, while they tended to ignore Japanese speeches with non-binary rhythms.

Furthermore, Kubozono (2008) showed that pitch accent in Tokyo Japanese is basically assigned on the basis of trochaic footing. Kitahara (2001) examined pitch accent location in Japanese based on the database by Amano and Kondo (1999): the results showed that accentual position of 50% of bimoraic frequently used words and 70% of non-frequent bimoraic words are word-initial, as well as most of the accentual positions in trimoraic words. These findings seem to suggest that the iambic pattern in Japanese is rarer than the trochaic. Rosen (2001) shows that trochaic words are present 40% more than iambic words in Yamato (ancient Japanese used around the sixth and seventh centuries). Iversen, Patel and Ohgushi (2008) also found that Japanese speakers tend to perceive a rhythm with a repetitive alternation of strong and weak sounds as a trochaic rhythm, and that the rhythmic perception of Japanese speakers was stronger than that of English speakers.

Another study on Japanese prosody by Port et al. (1987) supports the view that Japanese rhythm consists of bimoraic trochaic rhythm: they showed that, if a mora is shorter than average, preceding and following morae become longer than average. This anti-compensation denies that Japanese is a mora-timed language, while simultaneously showing that Japanese rhythm can be similar to English eurhythmy, with a trochaic alternation of prominent and non-prominent elements. However, due to the feature of pitch accent language, many Japanese content words do not have an accent. Considering all the sources cited in this section, it seems that, although the base rhythm or eurhythmy of Japanese is an alternation of accented and unaccented morae, the frequently occurring lapse hides the trochaic pattern. Moreover, despite the fact mora-timing has been denied by some studies, the Japanese mora structure is a simpler structure than the syllable structure of Germanic and Slavic languages. This means that the moraic duration can be considered relatively constant when compared to Germanic and Slavic languages. The regular pattern of syllable structure is one reason why Japanese was considered to be a mora-timed language. Additionally, the absence of schwa in Japanese is another possible reason early studies on timing suggested that Japanese is mora-timed, as, according to Beckman and Edwards (1994), durations of schwa and full vowels are different for the reason that vocal organs need more time to produce the sound when pronouncing full vowels. As there is no schwa in Japanese phonologically, the duration of mora can be relatively regular, and vowel quality does not work as correlates for Japanese pitch accent (Beckman, 1986).

Studies discussed in this section demonstrate the possibility that Japanese rhythm can be trochaic. Unfortunately, there is no study that indicates the frequent occurrence of lapse hiding basic rhythm or eurhythmy in Japanese. One problem is that most prior studies examine trochaic Japanese rhythm in limited situations, such as verse-like text, nicknames, or onomatopoeia, mostly based on intuitive perspectives. This means that early studies do not indicate whether the trochaic pattern is common in Japanese or if the rhythm is limited to specific types of word. If trochaic rhythm rarely occurs in Japanese, we cannot say that Japanese rhythm is trochaic. Another problem in early studies, on whether Japanese rhythm is trochaic and if Japanese pitch accent is rhythmically similar to English, is that they are based on intuition rather than on data and experiments.

To examine whether the Japanese rhythm is universally trochaic, a corpus study would be helpful in that it could quantitatively tell us the general tendency of Japanese rhythm that so far remains unexamined. The corpus is beneficial to solving an issue in rhythm study: the vastness of inter-speaker variability when identifying language-specific rhythm. It would also uncover whether Japanese pitch accent plays a role of prominence in rhythm formation similar to that in English. Japanese rhythm will be examined further in Chapter 4 and discussed based on the corpus.

2.2.3 Rhythm and accent of Japanese dialects

So far, accent and rhythm of standard Japanese have been discussed; however, many aspects of dialects, including accent and rhythm, are different from standard Japanese. This means that we need to recruit participants whose dialectic rhythm is identical. One of the major dialects in Japan is the Kinki dialect, which is spoken in the Kinki region, including Kyoto, Nara, and Osaka. The phonetic and phonological features of this dialect are summarised by Hyogaki (1962). According to him, vowels and consonants are slightly different from standard Japanese: for instance, /u/ is slightly more rounded than standard Japanese, and it is rare that vowels in Kinki dialects are devoiced, although they can be in standard Japanese (as mentioned in Section 2.1.). Consonants in Kinki dialects are almost identical to those of standard Japanese and, although the position of accent in a word is, in many cases, different from standard Japanese, the rhythmic feature of tolerating lapse is not. In addition, Hyogaki (1962) mentions that, similar to standard Japanese, a mora is a basic rhythmic unit in the Kinki dialect.

There is, however, a dialect in which the rhythm structure is different from standard Japanese and the Kinki dialect: according to Kubozono (2005), some dialects in the Tohoku region (North-eastern region, Akita dialect) and the southern part of Kyushu (Kagoshima dialect) are called syllabeme dialects. As the name suggests, a rhythmic unit in the syllabeme dialect is a syllable, while mora is the basic rhythmic unit in standard Japanese. For instance, がっこうしんぶん (school newspaper) [gak̚koɯʃiɴbɯɴ] is counted as four syllables in syllabeme dialect: gak̚/koɯ/ʃiɴ/bɯɴ. In standard Japanese and other Japanese dialects, these words are treated as eight morae, because the word is written with eight letters: ga/k̚/ko/ɯ/ʃi/ɴ/bɯ/ɴ. As the syllabeme dialects of Kagoshima and Akita have syllabic rhythm structure, Japanese speakers from these regions will not be appropriate participants for the current study. Kubozono (2005) mentions that syllabeme dialect is a syllable-timing dialect.

These rhythmic features in Japanese dialects must be considered in choosing participants for an experiment that examines rhythmic perception, which might be influenced by native language. Therefore, in the current study, native Japanese speakers from the regions of Kagoshima and Akita were not recruited.

2.3 Russian

2.3.1 Stress accent in Russian

Russian lexical stress is similar to English in some ways, for example, vowel reduction is found on unaccented syllables, and the accent is prominent while unaccented syllable is perceived as a non-prominent syllable as mentioned in the last section. Laver (1994) classifies both English and Russian as free-stress languages, which means that there is no fixed rule for assigning lexical accent on a word, while lexical stress position in fixed-stress languages, such as French (where the last syllable of a word is always accented) is fixed.

According to Jones and Ward (2010), Russian has five vowels (/i/ /u/ /e/ /o/ /a/) in stressed syllables and unstressed syllables are /ɐ/, /ə/, and /ɪ/. They also mention that Russian has a stress accent that works as the prominence in a rhythmic framework. Hyman (2006) also classifies Russian as a stress-accent language similar to English, adding that vowel reduction is a feature of an unaccented syllable in these languages. According to Crosswhite and Jun (2001), the vowel reduction in Russian means that unaccented syllables tend to approach schwa, being shorter with weak amplitude. As summarised by Barnes (2007), [e] and [o] appear only on a stressed syllable, while [i] and [u] appear on both stressed and unstressed syllables.

According to Bondarko (1998), Jones and Ward (1969), Kijak (2009), and Kondaurova and Francis (2008), vowel reduction is an important acoustic cue to stress-accent perception in Russian. Bondarko

(1981) and Svetozarova (1998) mention that Russian speakers rely on duration and intensity to grasp the stress accent, although intensity is a relatively weak cue compared to duration. The study on whether pitch is an important cue for Russian stress is limited, but Chrabaszcz, Winn, Lin, and Idsardi (2014) show that pitch is the least important cue for Russian speakers. Considering these studies on the cues for the Russian stress accent, it seems that duration and vowel quality both play important roles, while it is not clear whether pitch is a cue in detecting accent in Russian.

One feature of Russian syllabic duration might be the duration of the syllable preceding the stressed syllable (Bondarko, 1998). The duration of the pre-stress syllable is considered an approximate average of unaccented and accented syllables, as explained by Bondarko (1998). As is demonstrated by Chrabaszcz, Winn, Lin, and Idsardi (2014), the first syllable of ‘capaɸáH’ /sərɐ'fan/ (a traditional Russian dress) is shorter than the other two syllables, while duration of the pre-stress syllable /rɐ/ is shorter than the last stressed syllable /fan/, but longer than /sə/.

Padgett and Tabain (2003) show that there is no durational difference between unaccented syllables and pre-stress syllables, based on recorded data of nine Russian speakers, although the pre-stress syllable tended to be longer than an unaccented syllable. However, the inter-speaker difference was significant: two speakers’ pre-stress syllable durations were approximately 150% longer than an unaccented syllable.

2.3.2 Russian rhythm

Russian is a language with a stress accent similar to that of English; thus, Russian is also considered to be a stress-timed language (Abercrombie, 1967). However, in the current study (as is discussed in Section 1.3), linguistic rhythm is treated as a regular pattern of prominence and non-prominence.

Some studies, such as Alderete (2013) and Crosswhite, Alderete, Beasley, and Markman (2003), examined whether there is a default stress position in Russian, but no consensus was reached: Crosswhite et al. (2003) argued that the Russian default position of the stress is the first syllable of a word, but Alderete (2013) claimed that the position is the last of a stem. The number of studies regarding whether Russian tends to be trochaic or iambic is limited; however, Lavitskaya (2015) suggests that Russian is basically a trochaic-pattern language (prominent unit precedes a non-prominent one). In experiments by Lavitskaya (2015), 30 native Russian speakers read artificial words, designed specifically for the experiment to examine the position where native Russian speakers produce a stress accent: the results showed a strong trochaic tendency. This trochaic tendency is supported by Bethin (1998) who showed that Late Common Slavic language had a trochaic prosodic structure.

Russian has been regarded as a stress-timed language (Abercrombie, 1967), though Roach (1982) provides evidence that casts doubt on this assertion. He specifically examined the vocalic durations of three syllable-timed languages (French, Telugu, and Yoruba) and three stress-timed languages (English, Russian, and Arabic), and found that vowel duration did not differ more in the latter than it did in the former. Thus, the results of Roach (1982) do not support claims that Russian is a stress-timed language. Seemingly, it is doubtful that Russian is a stress-timed language; therefore, it would be necessary to define the Russian rhythm as a regular pattern of prominent and non-prominent elements.

The alternation of prominent and non-prominent rhythmic patterns in Russian is examined by Mills (1988): dialogue from six, native Russian speakers were recorded and then analysed. The results showed that, while the Russian stress accent is mobile (i.e. the accent position is not fixed), the actual stress accent position in Russian dialogue tended to be between unaccented syllables. This means that the Russian rhythm is an alternation of accented and unaccented syllables similar to the English rhythm discussed in Section 2.1.2. In addition to Mills (1988), Lavitskaya (2015) supports the regular alternation of prominent and non-prominent patterns in Russian, by showing that native Russian speakers tend to put an accent on that order in reading non-Russian words.

Moreover, Mills (1988) suggests that, in Russian, lapse and clash are avoided to maintain binary rhythm. Compared with Japanese, whose content words do not necessarily have an accented mora, lapses rarely occur in Russian, due to the accentual feature of the Russian stress accent that is found in all content words. As definite and indefinite articles are not used in Russian, function words that do not have an accent are fewer than in English, and there is no iambic tendency.

In Russian, Gouskova and Roon (2013) suggest that there is no secondary stress, except for a compound that only occurs in the following conditions (Gouskova, 2010):

1. Stress is on the left-hand stem.
2. The secondary stress is separated by at least two syllables from the primary stress.

The second condition seems to reflect a Russian rhythmic feature where a clash is considered an irregular rhythm and avoided. Although lapse is tolerated in the compound, it does not mean that it is tolerated in non-compound (Mills, 1988).

Through experiments in which Russian native speakers rated tolerance of accent position in a Russian compound, Gouskova and Roon (2013) found that Russian speakers do not tolerate clash in a compound and prefer the second stress to be far from the primary stress. Lapse constraint in Russian is not reported, and the preference of the lapse previously mentioned may suggest that lapse is tolerated in Russian.

Although the number of studies on Russian rhythm is limited, the studies mentioned so far show the basics of Russian rhythm clearly. In considering these studies, Russian rhythm can be summarised as follows:

1 Russian rhythm is an alternation of accented and unaccented syllables.

2 Lapse and clashes are avoided, which means these rhythms are treated as irregular rhythms, although lapse is tolerated more than clash.

3 The rhythm is a trochaic pattern (prominent followed by non-prominent).

Most importantly, these rhythmic features in Russian are identical to those of English, except for the preference of lapse in compounds.

2.4 Comparison between English, Japanese, and Russian rhythm

Looking at Section 2.1 and 2.3, English and Russian rhythms seem to be similar. However, according to the studies discussed in the previous sections, it seems that there are rhythmic differences between English and Russian, resulting from difference in cues to stress. As shown by Bondarko (1981) and Svetozarova (1998), duration seems to be a primary cue in Russian and both mention that loudness is a secondary cue. Chrabaszcz, Winn, Lin, and Idsardi (2014) show that pitch is the poorest cue in Russian, while it is also mentioned that vowel reduction is an important acoustic cue for Russian stress accent (Badanova, 2007; Bondarko, 1977, 1988; Janes & Ward, 1969; Kijak, 2009; Kondaurova & Francis, 2008; Kodzasov & Krinova, 2001; Zlatoustova,1953).

In English, duration seems to be the primary cue, while loudness is less important than the duration, the secondary cue is vowel quality, and pitch is the poorest cue for English stress accent (Kochanski, Grabe, Coleman & Rosner, 2005; Sluijter & van Heuven, 1996). Kochanski, Grabe, Coleman, and Rosner (2005) also show that pitch and loudness are important cues, while the pitch is the least reliable cue in English (see Section 1.2.2 for details of cues in English).

From studies on English and Russian cues, it is clear that duration is an important cue, while a pitch is the poorest cue in both languages. Comparing the rhythms examined in studies mentioned in sections 2.1.2 and 2.3.2, English and Russian seem to have a similar rhythm consisting of alternations of prominent and non-prominent syllables and avoiding clash and lapse. Another common rhythmic feature in English and Russian, is trochaic structure, whose prominence is expressed by stress accent. Lavitskaya (2015) argues that Russian rhythm is trochaic and Hayes (1995) argues that the English rhythm is trochaic. In conclusion, the eurhythmy of both languages is a trochaic alternation of stressed and unstressed syllables. The preference for lapse in Russian, as shown by Gouskova and Roon (2013), is a difference between English and Russian rhythm, although this preference is limited to only in compounds.

The discussions covered in this section so far suggest that English and Russian have different features of stress accents; however, the rhythm in both languages is an alternation of prominence and non-prominence.

These rhythmic characteristics of English and Russian seem to differentiate them from Japanese. As mentioned in Section 2.2.2, it is not clear whether Japanese rhythm consists of an alternation of prominent and non-prominent elements. In addition, it is not clear whether Japanese pitch accent can work as the foundation of Japanese rhythm, in a similar way to that of stress accent. Moreover, Japanese rhythm appears to be different from English and Russian as, due to frequent unaccented syllables, lapse (an irregular linguistic rhythm caused by successive unaccented syllables) is frequent in Japanese.

As many Japanese words do not have an accent (Tanaka & Kubozono, 1999), unaccented morae are frequently found in comparison to less-frequent accented morae, while all English and Russian content words have at least one accent.

In scrutinising the discussions mentioned in this section, if our native language affects rhythm perception, English and Russian should perceive rhythm similarly, due to the rhythmic similarity between these languages, although lapse is tolerated in Russian compounds. Due to the frequent use of lapse in Japanese, Japanese speakers may accurately perceive the non-binary rhythm that contains the lapse. Furthermore, the difference in primary cue to detect the accent may affect the results, depending on the prominence in the stimuli used in an experiment, as pitch is a primary cue in Japanese while it is the poorest cue in English and Russian. Although English and Russian have similar rhythmic structures, the preference in Russian might be a difference between them. The next section will discuss whether verse rhythms in English, Japanese and Russian are different from each other.

2.5 Metrics in English, Japanese and Russian literature and verse

In this section, metres in English, Japanese and Russian poetry are considered, as metrical structure is essential and reflects a language's rhythm. English and Russian verses are both classified as syllabo-tonic verse, which means that both numbers of syllables per line and the order of accented and unaccented syllables are controlled to make the rhythm regular (Scherr, 1980). Both English and Russian basic rhythms in verse are an alternation of a stressed and unstressed syllables. However, Scherr (1980) suggests that Russian words tend to consist of more syllables than English words, through comparing words in English works (*Moby Dick,* by Melville; *Huckleberry Finn,* by Twain; *The Sound and the Fury,* by Faulkner; and *A Farewell to Arms* by Hemingway) and Russian works (*Queen of Spades,* by Puskin; *Bela,* by Lermontov; *Dead Souls,* by Gogol; *A Nest of Gentry,* by Turgenev; *War and Peace,* by Tolstoy; *Crime and Punishment,* by Dostoevsky; *A Boring Story,* by Cexov; *The Petty Demon,* by Sologub; *The Temple of the Sun,* by Bunin; and *St. Petersburg,* by Belyj).

Table 2.2 Syllabic length in English and Russian by Scherr (1980)

Syllabic length of a word	English	Russian
1 syllable word	78.2%	15.49%
2 syllable word	17.15%	32.28%
3 syllable word	3.43%	28.44%
4 syllable word	1%	15.96%
5 syllable word	0.17%	5.94%
6 syllable word	0.04%	1.59%
7 syllable word	0.01%	0.3%

As shown in Table 2.2, Russian words tend to consist of more syllables than those in English. The data also showed that some English words had secondary stress while none of the Russian words did. As mentioned in Gouskova and Roon (2013), Russian secondary stress is found only in compounds. Scherr (1980) adds an example of a transliteration of rhythms in verse.

(2.21)

It was **bri**llig and the **sli**thy **to**ves
Did **gy**re and **gim**ble in the **wa**be

(2.22)

Var**ka**los’ **Xliv**kie sor’**ki**
Pyr**ja**lis’ po na**ve**

The examples above (2.21 and 2.22) show part of the original Lewis Carroll's Jabberwocky and its translation into Russian by Scherr (1980), in which, the bold letters suggest the accented syllables. Scherr (1980) mentions that the three or four syllables per line are accented in English while two or three syllables per line are accented in Russian.

These examples show that Russian rhythm tends to be more a lapse-like rhythm compared with English, due to a lack of secondary stress and long words. The idea from Scherr (1980) that lapse is more often found in Russian than English is also supported in a study by Gouskova and Roon (2013), which showed that lapse is tolerated in Russian compounds. For instance, kàrtəfʲilʲikʌpálkə (potato digger) is more tolerated by Russian speakers than kərtʌfʲèlʲikʌpálkə. This does not mean that English and Russian rhythms are different but, in looking at the rate of lapse, lapse would be more frequently found in Russian than in English.

Meanwhile English and Russian verses have a similar form, Japanese poetic forms called Haiku and Waka are different from English and Russian verse, similar to the difference between Japanese linguistic rhythm and English and Russian rhythm mentioned in the previous section. While each line of English and Russian verse consists of a constant number of syllables, numbers of morae in lines of Japanese verse are relatively irregular, for example, a haiku consists of three lines, but the length of each line is different, although the duration of reading each line is constant: the first and the last lines consist of five morae and the second line is seven morae. Tanka, another form of Japanese verse, consists of five lines and the number of morae per line is 5-7-5-7-7, respectively, while basic Japanese verse (haiku) consists of three lines and the number of morae per line is 5-7-5 as demonstrated in example (2.23) by Japanese poet, Basho.

(2.23) (bold letters suggest accented morae)

I.**mo.**u.e.te /iˈmouete/ (5 morae) ‘Putting potato’

Ka.do.wa.mu.gu.ra.no /ˈkadouɰamugulano/ (7 morae) ‘around there’

Wa.ka.ba.ka.na /ˈɰakabakana/ ‘young leaves’

The moraic number of each line is relatively irregular compared with English and Russian verse. Moreover, in Japanese poetic forms, the addition or deletion of a mora is tolerated. Here is an example of a haiku with an additional mora by Basho:

(2.24) (bold letters suggest accented morae)

Ta.**bi**.ni.**ya**.n.de /taˈbinijaNde/ (six morae) ‘Being sick while traveling’

Yu.**me**.wa.ka.re.no.o /juˈmeɰakalenoo/ (seven morae) ‘in dream, on a field’

Ka.ke.me.**gu**.ru /kakemeˈguru/ (five morae) ‘running’

A general interpretation of this piece is that Basho, a poet, was seriously sick in bed while traveling, but was running around a plain in his dream. As Japanese ‘n’, which is not followed by a vowel, is counted as a mora, the first line is counted as six morae, although the first line usually consists of five morae. This kind of addition and deletion of mora occurs freely in Japanese verse.

It may seem that Japanese verse is irregular in terms of rhythm, but Bekku (1977) states that Japanese speakers put pauses in reading the verse to arrange the rhythm, as shown in example (2.25).

(2.25)

μ μ - - -

μ μ μ -

μ μ μ - - -

In reading the verse shown in (2.25), a Japanese speaker makes eight units of rhythm per line by adding pauses, represented by hyphens in the example. A hyphen is similarly counted to a mora but is relayed by silence.

Based on an intuitive viewpoint, Sakano (1996) adds that, in reading a verse-like text, Japanese speakers tend to consider that a verse line consists of four bimoraic feet (as shown in the example below, although the last feet are silence). It is characteristic of this sort of Japanese literature that silence, or absence of voice, is also counted as a mora, and that the silence can correspond to a mora or three morae.

(2.26)

[μμ] [μμ] [μμ] [μ-]

Looking at the examples of Japanese verse, the rhythm seems to be similar to Japanese linguistic rhythm, in the sense that the regular rhythmic pattern is not straightforward. Even if the duration of the line is constant, due to the adjustment of inserted silence between lines, the duration number of morae per line is irregular (Sakano, 1996).

All examples in this section seemingly clarify the linguistic rhythm in English, Russian, and Japanese. Although the rhythms of Russian and English are similar, the Japanese rhythm is different from them even in verse.

Chapter 3: Musical Rhythm

3.1. Introduction

As is already mentioned in the previous chapter, the definition of rhythm in the current study is a pattern of accented and unaccented elements, or notes for music, which is identical to the metre in music. Therefore, a musical metre is synonymous with rhythm in the current study. This is because the metre in music suggests the repetitive pattern of accented and unaccented notes or beats. However, rhythm in music is used to refer, not only to the metre, but also to a durational pattern of each note (e.g. a durational pattern within a beat as shown in Figure 3.1). These seven metres are used in most types of music throughout the world.

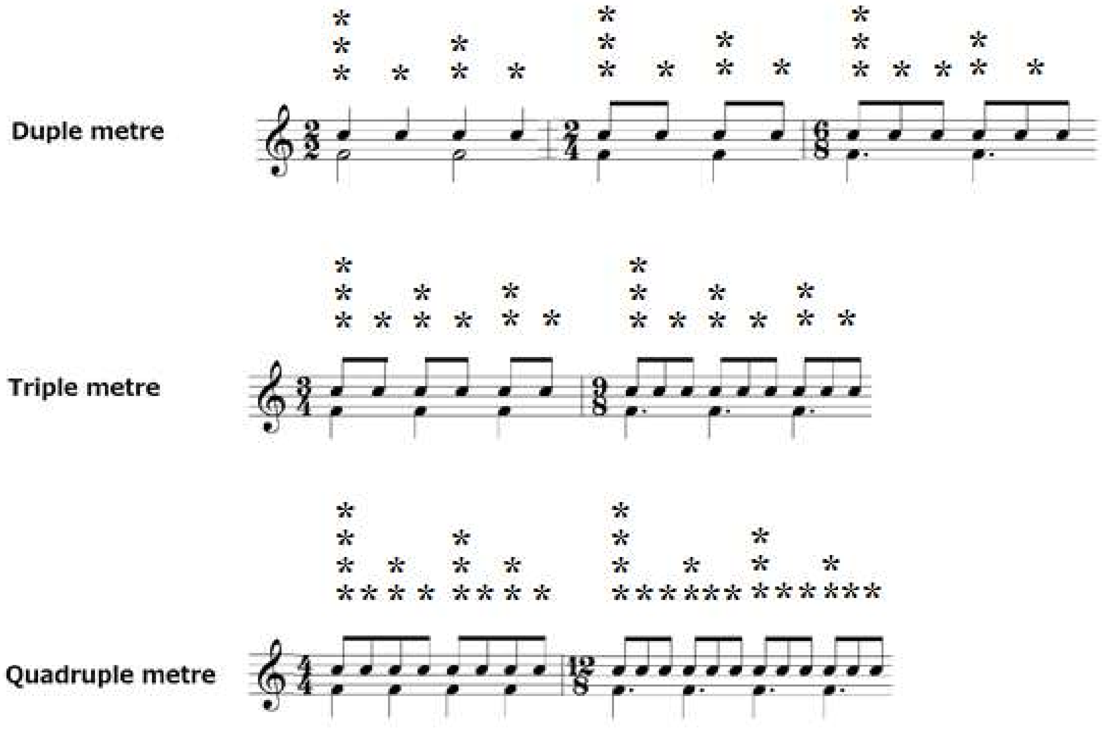


Figure 3.1: List of basic metres that are found universally regardless of a genre of music. The asterisks show the metrical structure.

The first staff of Figure 3.2 below shows examples of evenly divided rhythm, in which each note has identical duration. The second staff shows examples of unevenly divided rhythms. Although the patterns of prominent and non-prominent notes can be strictly regulated by metres, the structure within notes can be flexibly chosen in music.

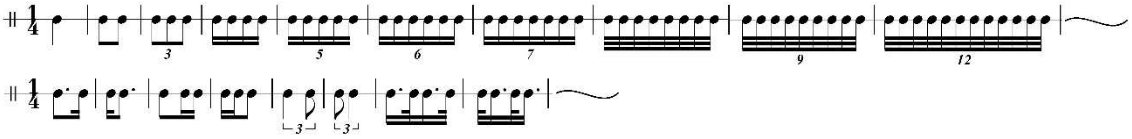


Figure 3.2: Examples of possible patterns within a beat.

In music, for instance, a metre 3/4 suggests that the first beat is an accented beat and the following two beats are unaccented, the pattern of these three beats are then repeated until the end of the piece. Basically, in most music types, the metre is unchangeable throughout the work. Usually duple, triple, and quadruple metres are used in most musical styles. Simple duple metres are, for instance, 2/4 and 2/2. Triple metres are 3/4 and 9/8. Lastly, quadruple metres are 4/4 and 12/8.

Even in classical music, or music for children (which usually consists of binary or ternary rhythm), rhythm or duration of each note can be improvised by performer or singer. Yet, the rhythm in most European music should be based on the notated notes in the score, which means that a strong violation of the metre is not tolerated; however, some folk music in Japan and Russian tolerate rhythmic irregularity (see Section 3.6).

Palmer and Krumhansl (1990) examined whether musicians and non-musicians perceive rhythmic violations differently. In their experiments, 20 musicians and 10 non-musicians were asked to compare two types of rhythm. One rhythm was constant, and the other rhythm had delayed or advanced downbeat, which made the duration of notes irregular. The participants rated the rhythmic difference between the two rhythms. Importantly, the word rhythm in the example by Palmer and Krumhansl (1990) means rhythm in musical meaning, defining that rhythm as a durational pattern of units in a phrase. They found that musicians could detect rhythmic violation of the delay or advancement of the downbeat better than non-musicians. In other words, the rhythmic violation of downbeat was easier to detect for musicians than non-musicians. This would suggest that the rhythmic perception of a musician is different to a non-musician. Also, Drake (1993) showed that musicians could reproduce musical rhythm better than non-musicians: after listening to drum-based stimuli, Drake's participants were asked to reproduce the rhythm by hitting a drum. Musicians could reproduce triple and complex metre better than non-musicians, while there was no significant difference in the reproduction of simple binary rhythms between musicians and non-musicians.

In this section, the possible perceptual differences between non-musicians and musicians have been addressed. However, it is also important to consider cultural music in England, Japan, and Russia and its cultural background. Folk music is formed as a result of cultural selection, which means that the understandable and preferable elements of music, such as form, rhythm, and scale, are historically and naturally selected and transmitted by people within the cultural group (Karpeles, 1951). Therefore, it is important to see the metres in folk music to consider the culture-specific rhythm. Also, by listening to the metres in folk music, it is possible to be sensitive to the rhythms used. The metres in English, Japanese, and Russian traditional music are discussed in the following sections.

Musical experience also affects rhythm perception (Hannon & Trehub, 2005). Therefore, musical education in England, Japan, and Russia will be discussed to take into consideration the outline of musical experience in these countries. Hannon and Trehub (2005) and Prieto (2012) both suggest that childhood is an important time to develop culture-specific rhythm perception; thus, it is necessary to consider music education in childhood in this chapter.

3.2 Examples of studies on the perceptual influence of musical experience

The studies by Cutler (2000) and Otake, Hatano, Cutler, and Mehler (1993) concerning the language of rhythmic or prosodic units (see Sections 1.1 and 1.4), suggest that adults count rhythmic structure differently depending on their native language. Some studies discussed in Chapter 2 showed that our rhythmic perception can be sensitive to a rhythm that is also used in the native language. Hannon and Trehub (2005) show that we are also sensitive to the familiar musical rhythm.

A study by Patel and Daniele (2003) shows a culture-specific connection between linguistic and musical rhythm, finding that musical rhythms used by English and French classical composers show similarities with the speech rhythm of their languages. They compared variability of syllabic duration of French and British speakers in Ramus (2002) with the variability of rhythms in musical works composed by the French and British composers listed in Table 3.1.

Table 3.1: Composers and their works examined by Patel and Daniele (2003)

Composer’s name	Native language	Example source
Arnold Bax	English	Sonata, Viola & Piano
Frederick Delius	English	Concerto, Violin & Orchestra
Edward Elgar	English	Concerto in B minor
Gustav Holst	English	The Planets
John Ireland	English	Sonata in G Minor
Ralph Vaughan Williams	English	A London Symphony
Claude Debussy	French	Les Parfums de La Nuit
Vincent d’Indy	French	Le Camp de Wallenstein
Gabriel Faure	French	Quartet in C minor
Arthur Honegger	French	Pastorale D’Été, Orchestra
Jacques Ibert	French	Concerto, Alto Sax
Darius Milhaud	French	Pastorale for Ob., Cl., Bsn.
Francis Poulenc	French	Toccato, Piano
Maurice Ravel	French	Introduction & Allegro
Albert Roussel	French	Symphony No. 3 in G minor
Camille Saint-Saëns	French	Concerto No. 3 in B minor

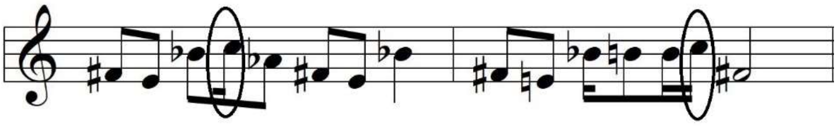
To avoid problems that can cause rhythmic biases, Patel and Daniele (2003) removed vocal music, which is inherently rhythmically influenced by its lyrics. They also focused on works composed in the twentieth century, considering that old linguistic rhythm, which could influence musical rhythm, would be different to the present. Finally, they excluded musical genres that could influence the rhythm, such as dance music and works for children. They showed that musical rhythm and linguistic rhythm in a country tends to be similar, for example, the duration of French linguistic and musical sounds tends to be less variable than those found in English. Even though they were careful of factors that could affect rhythm in musical works, there still might be problems. A metric nPVI, intended to show the durational variability of a vowel (see Section 1.3), was not very successful in analysing the linguistic rhythms (Arvaniti, 2009). What is necessary to consider is that Patel and Daniele (2003) used nPVI to compare musical works composed by English and French speakers, which means that problems caused by the metric would also be applicable in analysing durational variability in music. For instance, an inter-composer's differences can be so significant that the results of Patel and Daniele (2003) might not be reliable. Looking at the inter-composer differences and inter-work differences, although Patel and Daniele (2003) conclude that note duration in French music tends to be less variable than English music, the work of some French composers, such as works by Vincent d'Indy, were more rhythmic or rhythmically variable than English composers in their data.

Another problem in their study is that the metrical value could be musical-context dependent. Although they treated both solo instrumental work and orchestra work the same, their rhythmic structures are different, due to the fact that orchestra pieces must be written so that different instruments are synchronised, while solo work does not need to synchronise. This could mean that the rhythm in solo work could be more variable as there is no need for synchronisation.

Furthermore, they treated different instruments similarly; however, the agility of instrumental sound depends on the instrumental structure. It is clear that the way of producing piano sound is different from woodwind or brass instruments: a piece written for the horn cannot use the same rhythms used in piano solo work, because the horn cannot create complicated rhythms due to the unstable mechanism it uses to produce sound. The piano can play a role as a melody in an orchestra or chamber music, while the horn is supposed to support harmonic structure by producing long notes, making the horn rhythm less variable.

Even within string instruments, the rhythmic variability is different: for example, finger distance between each note on the double bass is wider than the violin, due to the size of the instrument, making it difficult to produce quick or short notes on the double bass. As a result, the note duration of the cello will be less variable than the violin.

A further problem is that musical tempo, which is the pace of a piece of music, was ignored in the study. A feature of musical tempo is its large variability, which can influence rhythm as, in performing rapid passages mainly consisting of sixteenth notes, it is impossible for a performer to express complicated rhythm. Therefore, it would be safer to compare musical works of a similar tempo. Another factor that we need to consider in comparing musical rhythm is that composers consider the harmonic structure in deciding rhythm (Piston, 1984), in other words, the musical works are not composed by intuition, but by consideration of a balance between tempo, form structure, harmonic structure, melodic structure, and rhythm. Taking into consideration all of these factors means that it would be difficult to see an influence of the composer's native language on their music, due to the impact of composer-specific considerations. In Example 1.1 (below) Olivier Messiaen, a French composer, intentionally inserts an additional sixteenth note (circled) for each measure, which makes the metre and rhythm more complicated.

(1.1) Example from 'Quartet for the End of Time' by Messiaen

Lastly, composers whose themes were analysed in Patel and Daniele (2003) were influenced by different or ancient musical styles: Fauré, Debussy, and Ravel all used modal harmonies used in rhythmically calm works (Howat, 2009); Debussy's works were influenced by Gamelan music (traditional Indonesian music) whose rhythm is constant (Mueller, 1986); and Ravel was born in the Basque region of France, close to the Spanish border, and his mother was of Spanish descent, so his music is partially influenced by Spanish music (Landormy & Wager, 1939). In addition, Ravel, Debussy, Poulenc, and Milhaud are all influenced by jazz music, whose musical rhythm is different from classical music.

The difficulties in considering culture-specific rhythm tells us the importance of examining ethnic or traditional music in a culture in a similar way to a linguistic corpus, through a database excluding individual intentions by composers. Subsequently, the rhythms in traditional music will be treated in this thesis. The problems listed above are similar to those of linguistic isochrony and metrics discussed in Section 1.3 (i.e. inter-composer differences and the reliability of metrics). This tells us that it would be advisable to treat the musical rhythm as a repetitive pattern of prominent and non-prominent units, similar to the definition of linguistic rhythm used in the current study.

Originally, the relationship between native language and non-linguistic sound perception was about pitch perception (Deutsch, Henthorn, & Dolson, 2004). Deutsch et al. (2004) claimed that both Vietnamese and Mandarin speakers whose native languages are tone languages performed better than English speakers in an experiment to judge absolute pitch (AP). AP is a rare ability that enables us to identify the pitch or to produce indicated pitch correctly without a reference tone. They discuss that tone-language speakers tend to have absolute pitch, as the meaning of words in their language differs depending on pitch. For instance, when AP holders listen to a short phrase consisting of C, D, and E tones, they can easily identify the tones used in the phrase; however, relative pitch (RP) holders (people who do not possess absolute pitch), it is impossible to identify pitch without a reference note. In other words, RP holders can only identify the phrase after being notified of the tone of the first sound in the phrase. RP holders can then identify the pitches of other tones by judging the interval between C and D, and C and E. Furthermore, Deutsch, Henthorn, Marvin, and Xu (2006) show that early musical education and exposure to Mandarin (a tone language) help to acquire absolute pitch, possibly because pitch plays important role in distinguishing meaning in the tone languages.

However, Gregersen, Kowalsky, Kohn, and Marvin (2001) suggest that early childhood music education and genes can affect the acquisition of AP through data showing that many Asians, 26% of Japanese, and 37% Korean speakers, whose native languages are not tone language, are also AP holders, compared to only 9% of Caucasian students. It may be understandable that many Japanese speakers are AP holders because pitch plays an important role in distinguishing Japanese lexical meaning in a similar way to tone languages; however, from the perspective that native language affects AP, it is not clear why many Korean speakers, whose native language does not include pitch discrimination, are AP holders. According to Jun, Kim, Lee, and Jun (2006), in one of the Korean dialects (Northern Kyungsang Korean), the position of pitch accent can change lexical meaning similarly to Japanese, but this does not explain why the percentage of AP holder in the Korean group was higher than that of the Japanese. Ramsey (1991) shows that the Proto-Korean pitch accent was distinctive, but such an accent is not used in modern standard Korean.

Henthorn and Deutsch (2007) argue that Asian musicians who spent their early childhood in Asia had a higher prevalence of AP, compared to Asian musicians who spent their early childhood on the American continent. This suggests that native language can affect pitch perception, but the number of participants in the study was rather small (six to seven participants per group of Asians who spent early childhood in the American continent). Also, Henthorn and Deutsch (2007) claim that, since China, Japan, and Korea are countries of early music education, this early education affected the high prevalence of AP in Asian groups. However, they provide no source or evidence that shows this early music education in Asian countries. As is mentioned by Krumhansl (1991), early music education influences the acquisition of AP, and this might explain why the results obtained by Deutsch et al. (2006) showed more Mandarin speakers to have AP compared to other non-tonal language speakers: it was not because of the influence of their native language, but due to early music education in Asian countries (Henthorn & Deutsch, 2007). Moreover, even the reanalysis of data by Henthorn and Deutsch (2007) – in which they assess only Asians who spent their early childhood in Asian countries – found that 10/19 Koreans, whose language does not distinguish pitch, were AP holders, while only 3/7 Japanese, whose native language makes lexical distinctions based on pitch, were AP holders. There is no explanation for this high prevalence of AP in the Korean group in Henthorn and Deutsch (2007). Baharloo, Risch, Gitschier, and Freimer (2000) suggest the strong influence of genetics: by comparing the siblings of AP holders with a control sample, they found that siblings of AP holders tend to also be AP holders.

Looking back at Deutsch et al (2004) and Deutsch et al. (2006), it is not convincing that speaking tone language boosts the possibility of processing AP, solely because pitch plays an important role in the tone languages. Although not previously discussed, what tone language speakers do when listening to different tones is judge the pitch of each syllable, comparing them with the pitch of other syllables in a context. This means that the process of tonal language is identical to what relative pitch (RP) holders do when judging pitch in musical sounds. Therefore, it is reasonable for tone language to affect RP perception, but it seems there is no stronger connection between tone language and AP than there is between tone language and RP. Moreover, it seems reasonable that genetic factors affect AP more than a person's native language.

3.3 Musical rhythms and education in England

In this section, folk music in English is discussed in considering rhythmic perception and culture-specific rhythm. In this study, England literally suggests the specific part of the United Kingdom. According to MacKinnon (1993), who distinguishes English music from Irish or Scottish music, specific folk music in England includes the hornpipe, jig, Morris dancing, and sea shanties. Some may regard the ballad as traditionally English, but it is advisable to exclude it as the ballad was also developed and popular in other European countries, meaning that this genre is not a characteristic musical form of England (Ling, 1997). However, the term 'ballad' can refer to different genres, depending on the country. Some may consider the carol to be English music, but, according to Studwell (1995), the carol has also come from many cultures.

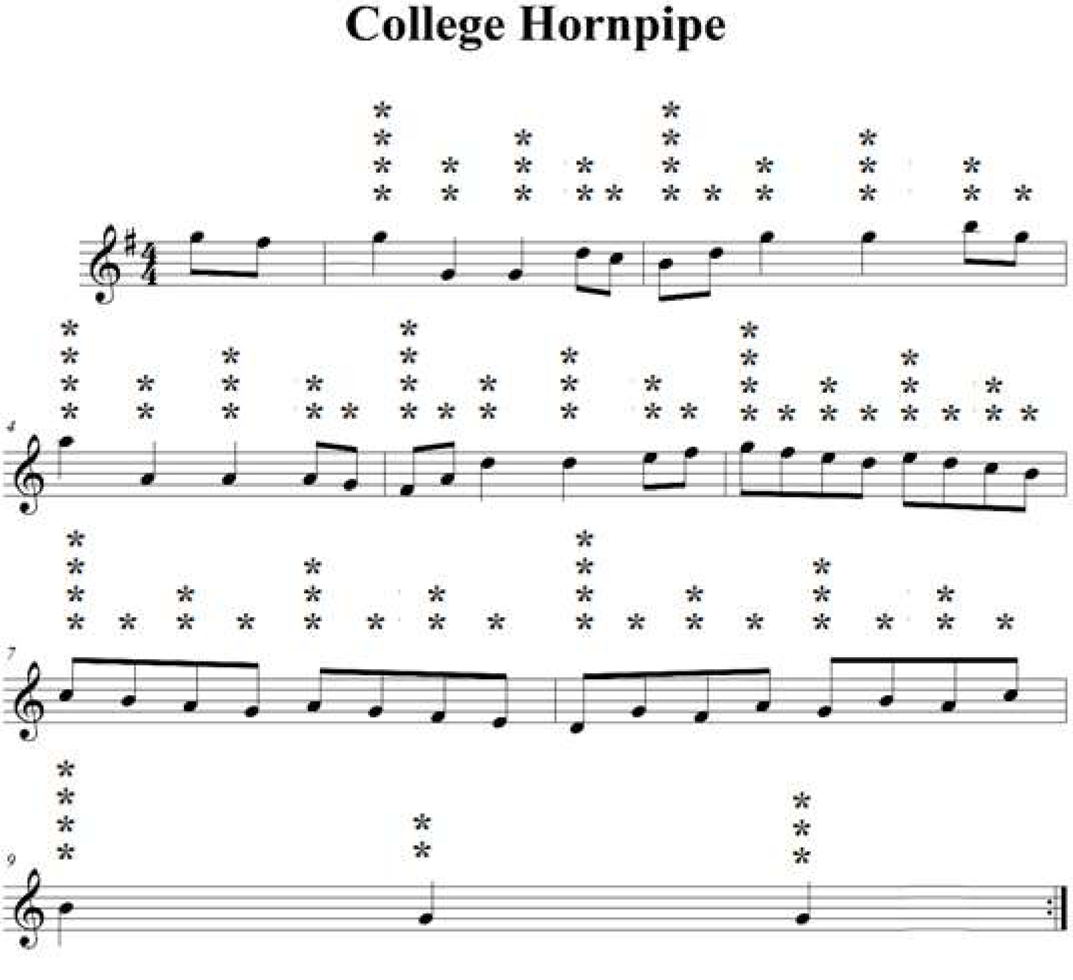


Figure 3.3: A score of well-known hornpipe work, 'College Hornpipe', with a metre of 4/4. The score is notated by Finale 2010. The asterisks show metrical structure.

Although there is no consensus on the origin of the hornpipe, it is generally believed that it was already prevalent around the sixteenth century. Hornpipe is an English, Irish, and Scottish sailors' dance. Figure 3.3 (above) shows the typical hornpipe rhythm and metre. The metres of this dance music, 2/2, 3/2, and 4/4 and their variants, are basic dance metres found in many other types of dance music.

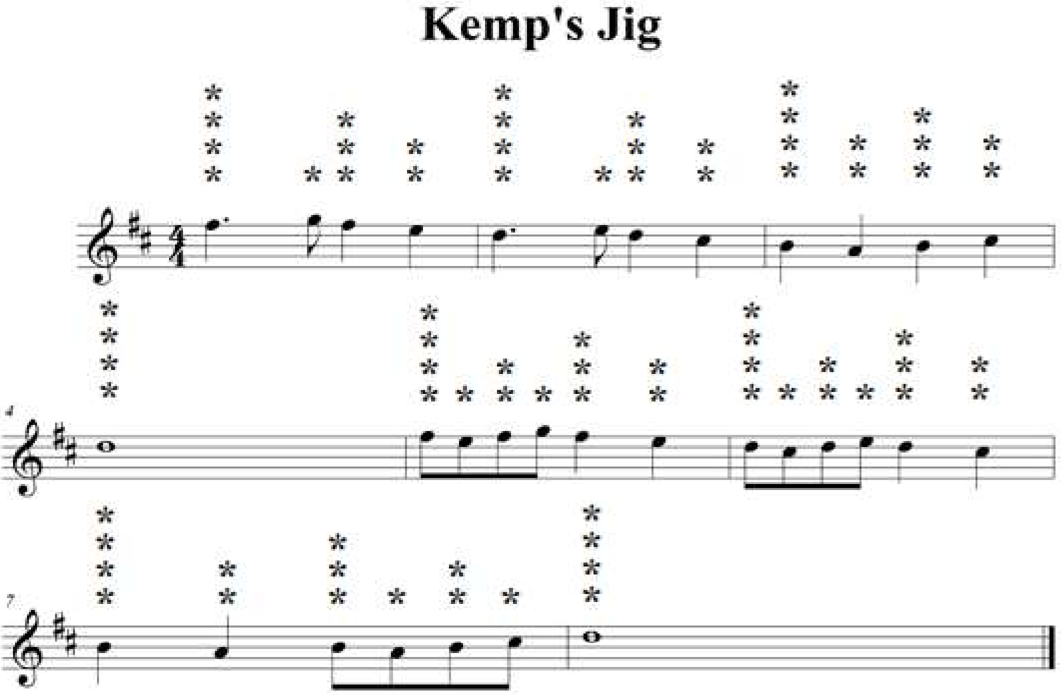


Figure 3.4: A score of jig work, 'Kemp's Jig', with a metre of 4/4. The score is notated by Finale 2010. The asterisks show metrical structure.

The jig is regarded as a dance from the sixteenth century and was developed in England, Ireland, and Scotland. A characteristic metre of this dance was the compound metre 12/8, which is a variant of the basic metre 6/8. However, this compound metre was gradually simplified. As a result, the metre of a jig today is 6/8. Figure 3.4 (above) shows the typical jig rhythm and metre.

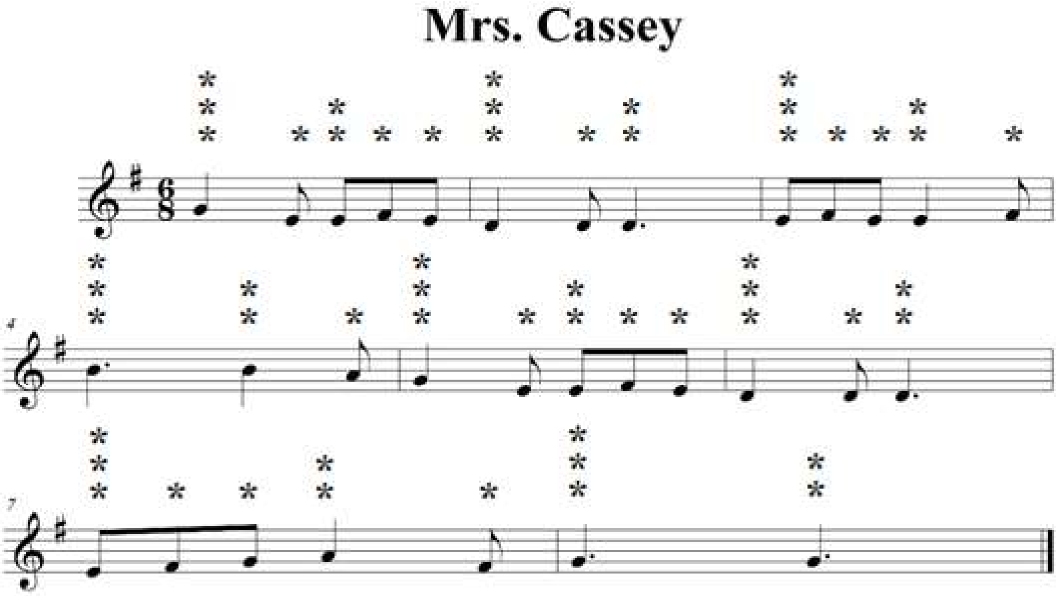


Figure 3.5: A score of a Morris dance, 'Mrs. Cassey', with a metre of 6/8. The score is notated by Finale 2010. The asterisks show metrical structure.

Morris dance is considered a form of English folk dance. The first description of Morris dance dates back to the fifteenth century. There are some sub-divisions of this genre, such as Border Morris and Molly Dancing. In all styles of Morris dance, metres are 6/8, 9/8, or 4/4. Figure 3.5 (above) shows the typical Morris dance rhythm and metre.

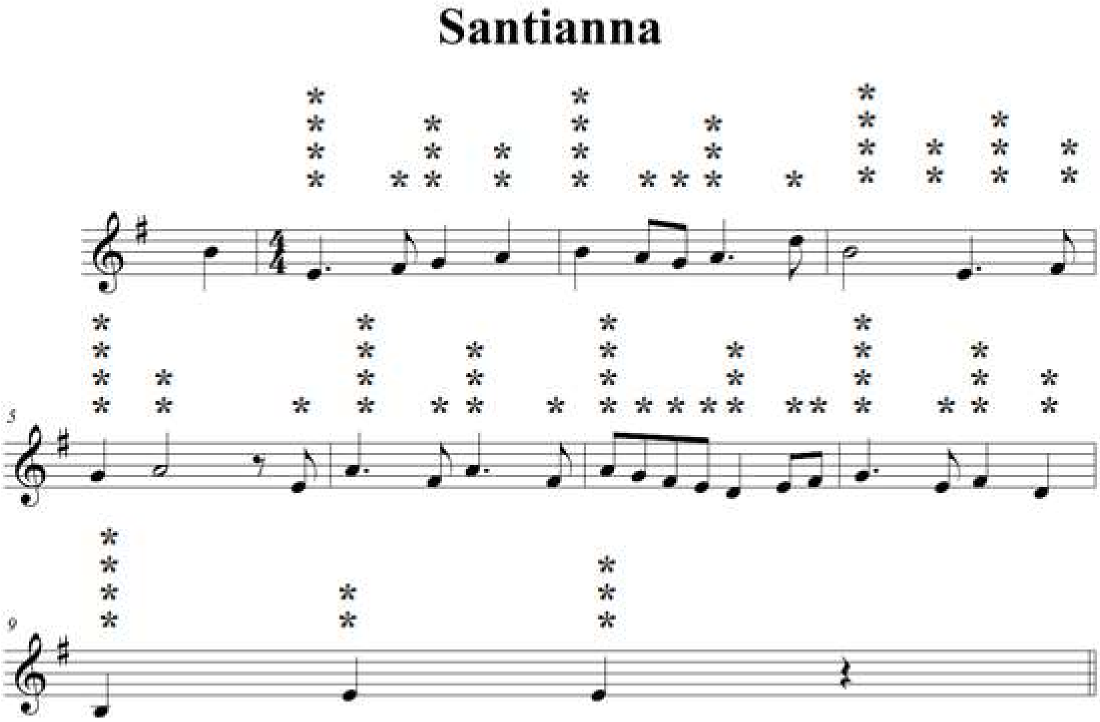


Figure 3.6: A score of sea shanty work, ‘Santianna’, with a metre of 4/4. The score is notated by Finale 2010. The asterisks show metrical structure.

Finally, sea shanty is a work song of sailors. This song was sung by sailors on merchant sailing vessels around the fifteenth century. The origin of the sea shanty is a British work chant, ‘shanty’. This type of work song was used to coordinate the movement of workers. The metres used in the sea shanties are mostly 4/4 or 6/8. Figure 3.6 (above) shows the typical sea shanty rhythm and metre.

Looking at the four genres of folk music in England, it is clear that all metres are simple duple, triple, or quadruple metres. Like English music, metres in most western European music are also basic duple, triple, and quadruple metres. It is also clear that duple metres are common in England, which may imply a correlation between the English musical metre and the quadrisyllabic rule proposed by Hayes (1984). The rule suggests that, when a distance of metrically most prominent syllables are spaced four syllables apart, the metre is eurhythmic. As the most prominent sounds in the duple metre are usually four beats apart from each other, it seems that the quadrisyllabic rule is correct in both English language and music.

Figure 3.7 suggests the metrical grid of musical structure and stress accents in lyrics. It shows that the musically prominent notes, expressed by three or two asterisks, are all assigned to accented syllables.

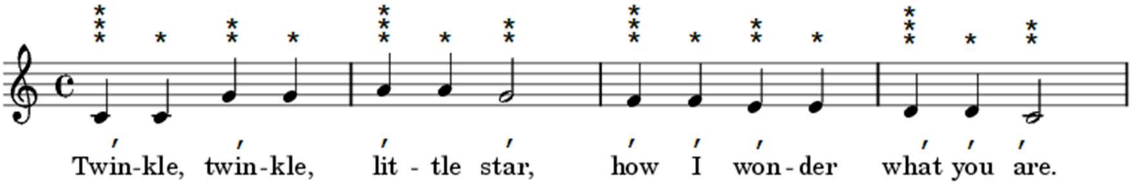



Figure 3.7: Prominent and non-prominent pattern written by asterisks for musical prominence and acute accent for a lyric.

Some may argue that triple metre, which is also used in English music, is out of the quadrisyllabic rule, as the prominent notes are spaced three syllables apart. However, considering the two musical terms phrase and sentence, which suggest a unit of musical metre, the most prominent note in the triple metre can be considered as four units apart from each other. This interpretation is based on the most prominent note in a measure being the first note and that the phrase or sentence in a musical piece consists of four measures. Therefore, the most prominent note of a sentence and phrase is spaced four measures apart. For instance, in Figure 3.7, the entire staff, which consists of four measures, is a musical phrase consisting of four units (prominences).

From the examples of English music and English linguistic rhythm discussed in this section, it is evident that English speakers are unfamiliar with non-binary rhythm, both in their language and music. Considering the definition of the rhythm in this study is the pattern of prominent and non-prominent sound elements, it means that binary rhythm suggests that prominent and non-prominent elements alternate in a binary pattern, while non-binary rhythm suggests that prominent and non-prominent sounds occur randomly (see Section 1.1 for the definition of rhythm). However, it is necessary to consider music education in England, which may increase sensitivity to the binary rhythm.

The British government issues a national curriculum for music. (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/239037/PRIMARY_national_curriculum_-_Music.pdf). According to the curriculum, the government considers the teaching of folk music important. A governmental guideline titled 'Designing and timetabling the primary curriculum' shows desirable teaching hours: 30 hours of music lessons per year is advisable for key stage 1 and 33 hours for key stage 2. In the curriculum, there is no detailed description of traditional music, which means music education for traditional music would vary from teacher to teacher.

Table 3.2: Teaching hours of music in England

Age	Total teaching hours of music	Teaching hours for English traditional music
5	30 hours	No description
6	30 hours	No description
7	33 hours	No description
8	33 hours	No description
9	33 hours	No description
10	33 hours	No description
11	33 hours	No description
12	33 hours	No description

Table 3.2 shows that students in England study music from 30 to 33 hours, per year, at school. These teaching hours are less than in Japanese and Russian schools (see Sections 1.3 and 1.4).

As is mentioned already, English classical, folk, and pop music largely uses regular binary rhythms. Thus, familiarity with English music and music education would be beneficial to the perception of binary rhythm. Although some progressive English rock music, like that of Peter Gabriel (Temperley, 2018) uses irregular meters, exposure to such works would be relatively rare for infants and children. According to Hannon and Trehub (2005), rhythmic perception is influenced by environmental musical rhythm mainly in childhood; thus, the infrequent rhythm in British music will not affect perception.

In conclusion, English musical rhythm tends to be binary rhythm, while non-binary rhythm is significantly rare. This is helpful in considering the influence of linguistic experience on rhythm perception, as an English speaker's musical background is totally different from a Russian's (see Section 3.4), while linguistic rhythmic features in English and Russian are similar. If there is a rhythmic perceptual difference between English and Russian speakers, it would not be due to musical experience, but due to linguistic experience.

3.4 Musical rhythms and education in Japan

In Japan, metres in many pieces of folk music are non-binary. There is a worker's song called 'Tanko Bushi' (which roughly translates to 'coal mine phrase'), which was sung by coal mine workers while digging.

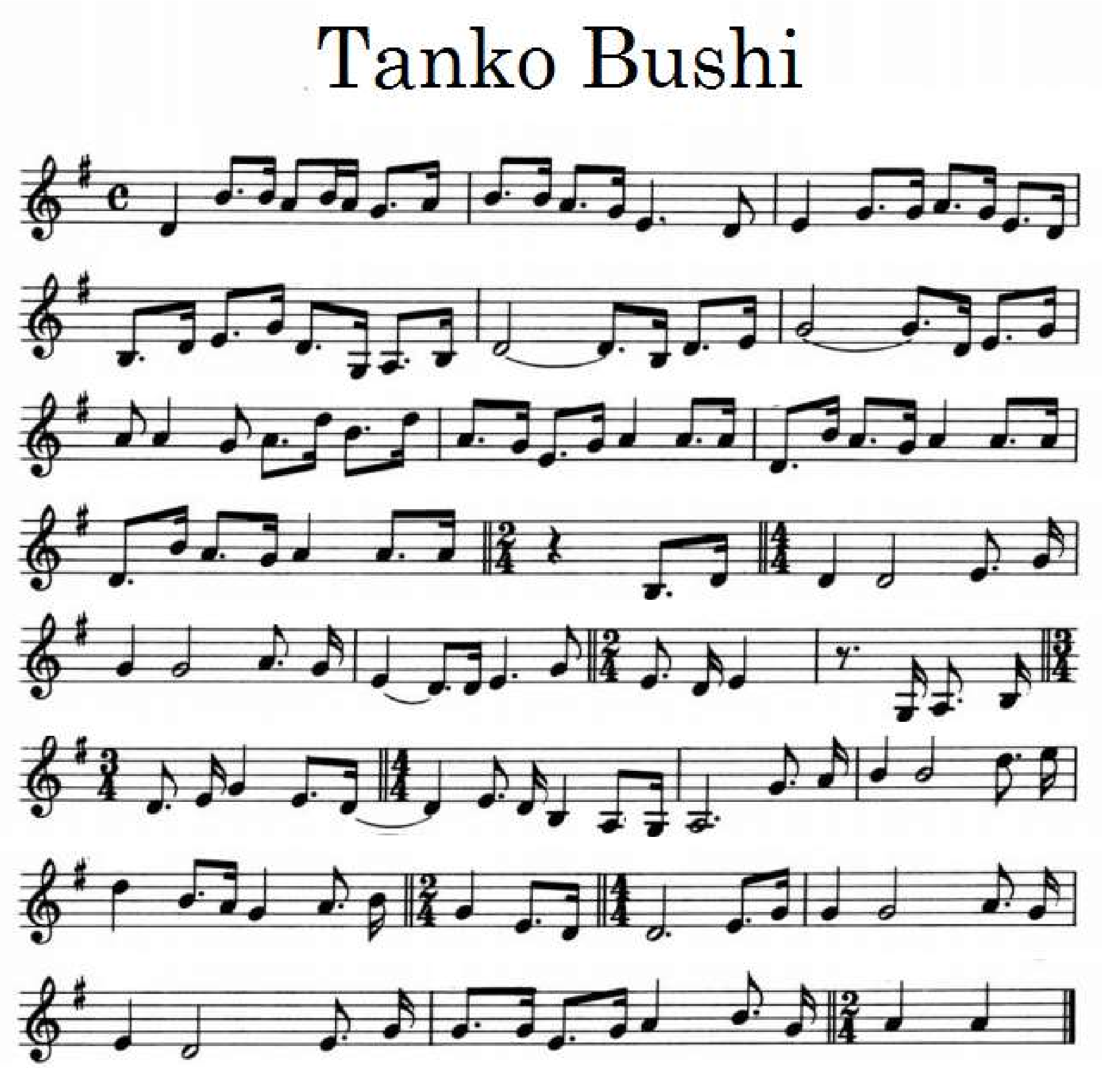



Figure 3.8: A score of traditional Japanese dance music, ‘Tanko Bushi’, with a frequently changing metre.

According to the musical score in Figure 3.8, the metre of ‘Tanko Bushi’ frequently switches from 4/4 to 2/4 or 3/4. Such frequent switching in metre is exceptional in western music, even in East European music, which tolerates compound metre such as 5/4 (2+3/4). In some sheets, editors use different time signatures for ‘Tanko Bushi’, and there is no consensus on the metre of this work. Nonetheless, surprisingly, this complicated rhythm was originally sung by workers to synchronise their movements, and the work is now used as dance music for Bon-Odori, a traditional dance to honour the spirits of ancestors. Many summer festivals hold the dance event and people join the dance freely. Although the original purpose of the work was to synchronise workers' movements, in a similar way to the sea shanty in England, the metre in the Japanese song is more complicated.

Complicated metres are also used in Japanese children's songs. A notable example of this would be 'Antagata Dokosa' (From Where You Are), whose metre changes frequently.

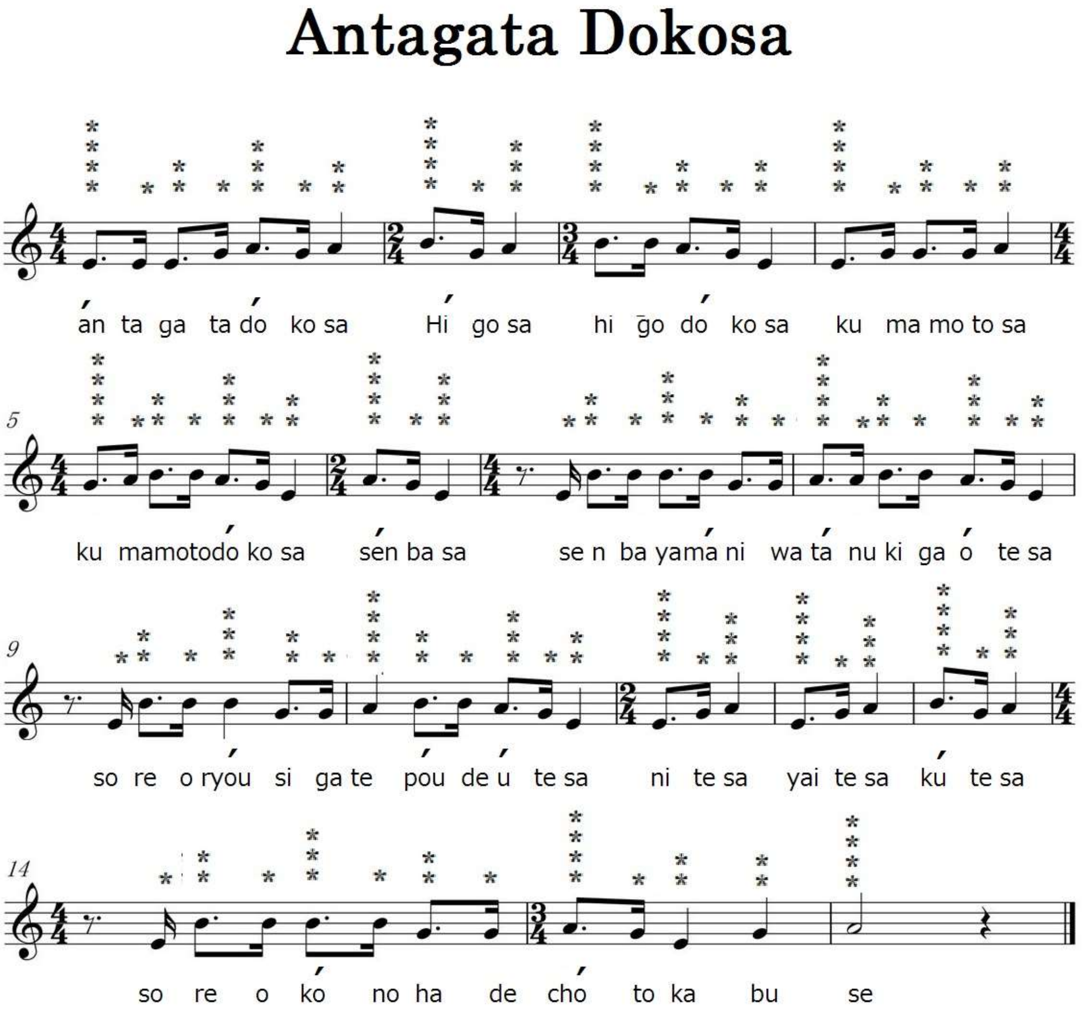


Figure 3.9: A score of a Japanese children's song, 'Antagata Dokosa', with a frequently changing metre. The asterisks show musical metrical structure while acute accent marks (´) on the text suggest pitch accent.

As illustrated in Figure 3.9, it is clear that the musical metrical structure is not constant, compared to examples of English music (see Section 3.2).

Another feature of Japanese music is that the pitch accent is not necessarily musically accented. For example, in measure 5 of Figure 3.9, the first note is a musically prominent sound (as shown by the four asterisks); however, the corresponding text 'ku' is unaccented mora, while an accented syllable 'ma' in measure 7 is in a musically unaccented position. In Japanese songs, such mismatch of musical and linguistic accent frequently occurs, as lapse, an irregular rhythm caused by successive unaccented syllables, frequently occurs in Japanese (see Section 2.2). Contrary to this, looking at the last section, the stress accent of English was also musically accented by downbeat, and unstressed syllables in English songs are usually musically unaccented due to upbeat. The examples in this section (a Japanese work song and children's song) show that change of metre was tolerated in Japanese culture and this tendency might influence linguistic rhythm in Japanese.

Another example of a non-binary rhythm in Japanese music is found in Japanese traditional music, 'Gagaku' and 'Noh'. Looking at Figure 3.10, it is clear that the rhythm is irregular, due to lack of constant rhythmic units (i.e. an apparent beat is not used). Moreover, there is no notion of a metre, measure, or bar. A difference of the non-binary rhythm in Japanese traditional music, from the irregularity of metre in the previous examples, is that there is no beat – neither downbeat nor upbeat, which works as a rhythmic unit in some forms of Gagaku and Noh – which means that there is no metre. Clayton (1996) and Nelson (2017) both treat the irregularity of Japanese rhythm in traditional music as 'free rhythm'. In music, the duration of a beat is generally constant throughout the work and, more importantly, even when the metre changes, the duration of beats does not change, as it is impossible to synchronise without a constant beat, although synchronisation in some Japanese music is ignored. As the duration of each sound in Gagaku or Noh music is freely decided by the performer or singer ad-lib, a metre that consists of accented and unaccented beats cannot exist. This rhythmic feature makes it impossible to notate notes in traditional Japanese music using western notation; in Japanese notation, the duration of each note and metre are not written.

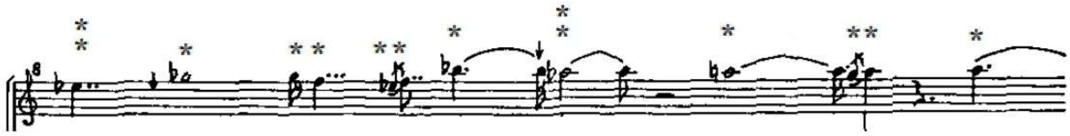


Figure 3.10: A score of a Gagaku work, 'Katen no Kyuu, showing an East-Asian-specific rhythm structure. The example is cited from Kubota (1969). It is impossible to describe metrical grid for the rhythm here, as there is no notion of metre in this type of music.

Figure 3.10 (above) shows the rhythm in 'Katen', a Gagaku work, which was notated based on recorded performance. As performers can decide the rhythm ad-lib, the notes written in Example 8 seems to be irregular. The example of Katen is notated in western notation style, but it is evident that the rhythm is more complicated than English music. Also, it is worth mentioning that the time signature, which describes the metre, is not featured in the example, as there is no fixed metre in the work. Some works in Gagaku and Noh consist of fixed metre, but even then, such metres can be irregular, for example, 8/4 or 5(2+3)/4.

Although the examples above are those of non-binary rhythm, simple duple, triple, and quadruple metres, such as 2/4, 3/4, and 4/4, are also used in Japanese folk music (Honda, 2014). Looking at examples of Japanese music, it is clear that it tolerates complicated metre (such as 8/4 and 5/4), changeable metres, and the absence of metre altogether. This does not necessarily mean that Japanese music generally consists of non-binary rhythms. In fact, rhythms in most Japanese works consist of simple metres such as 3/4, 4/4, and 6/8. Furthermore, although lapses are found in the examples provided (8/4 and 5/4), a clash is not tolerated in Japanese music.

Although there is an exception in the children's song, Angatata Dokosa, which is more popular than the aristocratic music of Noh and Gagaku, as it does not use the irregular duration of notes used in Noh and Gagaku. Instead, even in its mixed metres, the metres tend to be an alternation of prominent and non-prominent notes similar to Japanese language rhythm, and lapse is also found in Figure 3.9, which is also consistent with Japanese language rhythm.

In Japan, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) issues curriculum guidance in which class hours for music are strictly decided for key stages 1 to 4. The guidance states that all students have to study Japanese traditional music, including Gagaku and Noh. All teachers are supposed to submit a plan of the class schedule for each subject. After the submission, a headteacher checks whether the scheduled plan is appropriate, based on governmental curriculum guidance. Through this process, musical education in schools is standardised in Japan. This system is almost identical to Russian music education (Section 3.4).

Table 3.3: Total teaching hours of music in Japan

Age	Total teaching hours of music	Teaching hours for Japanese traditional music
5	No data (Infant school)	No description
6	68 hours	No description
7	70 hours	No description
8	60 hours	No description
9	60 hours	No description
10	50 hours	No description
11	50 hours	No description
12	45 hours	No description
13	35 hours	No description
14	35 hours	No description

Table 3.3 shows that teaching hours of music in Japan gradually decrease with an increase in age. According to the guidance, there is no description of teaching hours for traditional music. Nonetheless, it is written in the guidance that all pupils of key stage 1 have to study and listen to Japanese traditional music, and during key stage 3 (11–14 years old), pupils have to perform an instrumental piece of traditional music, using koto (harp-like instrument), shamisen (plucked string instrument similar to a guitar), shakuhachi (woodwind instrument), etc.

The total teaching hours in a Japanese infant school (between the ages of three and five years) are 156 hours per year. Teachers in infant schools are supposed to teach five subjects: gymnastics, language, communication, social studies, and arts (drawing and music). Each teacher can decide the schedule freely, which means the detailed plan for each subject is not provided by the MEXT. Unfortunately, there is no data covering the teaching hours for music in Japanese infant schools, as the curriculum is significantly different from school to school and teacher to teacher. Education in infant schools is not compulsory, but, looking at statistics offered by the Ministry of Health, Labour, and Welfare (MHLW), in 2016, 87% of three-year-olds, 97% of four-year-olds, and 98% of five-year-olds attend infant school. In Japan, children between the ages of six and twelve receive compulsory education in primary school.

From this section, it is clear that Japanese people are familiar with non-binary rhythm, due to their traditional music, dance music, and children’s songs. In addition, music education in Japan focuses on Japanese traditional music, whose rhythm and metre can be non-binary.

3.5 Musical rhythms and education in Russia

As mentioned in Section 2.5, Russian rhythm is trochaic similar to that of English rhythm, but the rhythms of Russian folk music are different from those found in English folk music. While musical rhythms in English folk music tend to be constant (duple, triple, or quadruple metre throughout the work), there are many Russian folk pieces (or tunes) with frequently changing metres.

Traditional Russian music has non-binary rhythm (irregular metre), which is different from the rhythms found in Western European music traditions. The complex metre is usually a successive alternation of two different types of meter: for instance, the alternation of a quadruple and triple metre is repeated constantly throughout the work, which consists of a complex metre. Some Russian folk songs have a complex metre, as shown in Figure 3.11. The metre in Figure 3.11 is 7/4 (3+4/4), in which the first and fourth beats are musically accented. Looking at Example 9, it is clear that linguistic accents (stressed syllables) are sung musically strongly. Example 9 also seems to show the Russian trochaic tendency.

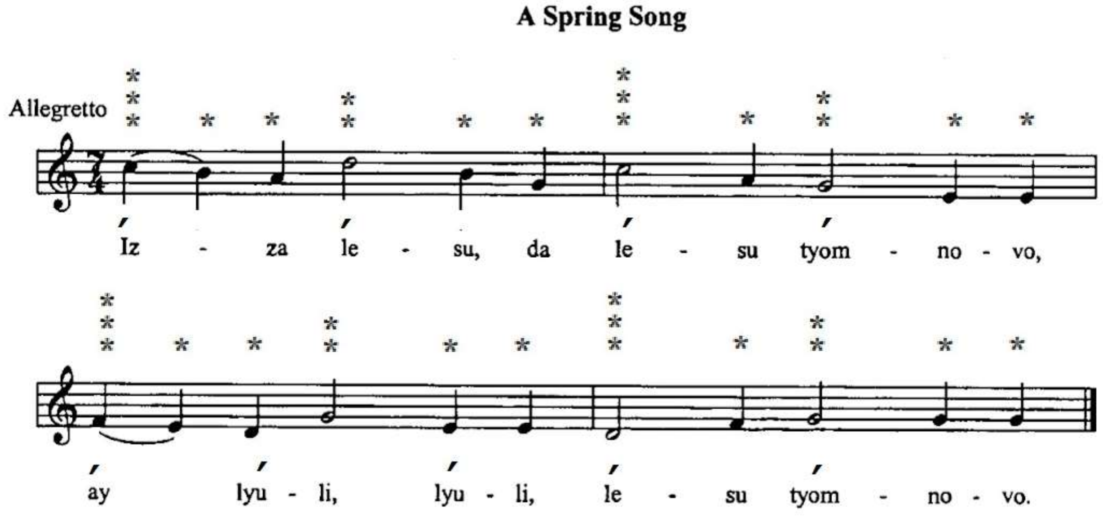


Figure 3.11: A score with a complex metre (3/4 + 4/4) from a Russian folk song from Prokhorov (2002). Acute accents suggest stressed syllables and the asterisks show musically prominent and non-prominent beats.

However, Russian folk music also has different non-binary rhythms from a complex meter, in the sense that the metre in Russian folk music can change frequently and randomly within a piece, like that shown in Figure 3.12.

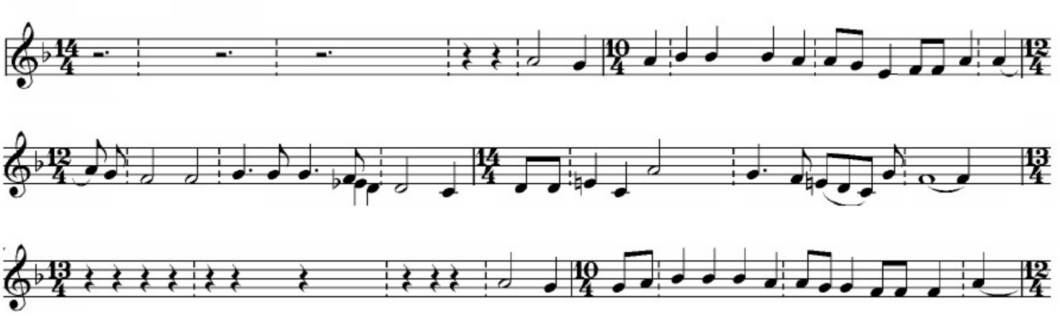



Figure 3.12: A score of an example from 'Скакал казак через долину' ('The Cossack was crossing the valley'). The score is cited from 'Сборник казачьих песен Белгородской области' (Collection of Cossack songs of Belgorod region) (2014).

Another difference between Russian musical rhythm and that of the English, is that the duration of some notes in Russian folk music is free, meaning a performer can change the duration of each note ad lib. (Prokhorov, 2002). Rhythm or duration of any notes can be free by speeding up or slowing down at the discretion of a performer, even in classical music, which consists of regular rhythm. However, the rhythm in most European music should be based on notations in the score. As performers in Russian folk music can change the rhythm freely, ad libitum, many Russian folk pieces are not written as musical scores – instead, the works are taught by ear, without a written medium. Subsequently, there is no strict rhythmic structure in Russian folk songs. Although some other European music has complex metres, such as 4+3/4 in Balkan music, the free rhythm and frequently changeable metre are characteristics of Russian folk music; these irregularities are also found in Asian music, such as Japanese music, rather than in European music.

Internationally renowned examples of Russian music with non-binary rhythm, are songs sung by Olga Sergeeva (Ольга Сергеева), whose recording was used in the film *Nostalgia* (ностальгия), directed by Andrei Tarkovsky. From her recordings, it is clear that metre, rhythm, and duration of each note are irregular and continuously repeat a phrase. As the rhythmic structures are too complicated to understand, it is reasonable that musical phrases are repeated until a listener understands them. In her repertoire, there is no binary rhythm; however, it is impossible to show the rhythm visually through notation. Some works appear similar to 3/4 or 4/4, but they do not perfectly fit with such regular metres. Recently, many pieces of traditional music or folk music are arranged in the style of pop-genre; however, her songs are truly Russian folk music. This sort of folk song is usually transmitted orally, rather than notated in a score. Pelageya Khanova (Пелагеея Телéгина) and Evgenia Smolyaninova (Евгения Валерьевна Смольянинова) are also known as Russian folk-song singers. In some of the works in their repertoires, non-binary metres can be found, but such metres are uncommon, due to their pop-style. However, songs performed by Pelageya Khanova and Evgenia Smolyaninova are broadcast on TV and radio in Russia, enabling listeners to become accustomed to the non-binary rhythms found in folk songs.

Other examples of non-binary rhythms in Russian music are found in works sung by Cossack males. Cossacks are a group of predominantly East Slavic speakers who were members of semi-military communities. They have their own culture and music. While previous examples of a non-binary metre in Russian music are found in songs sung mainly by solo female singers, Cossack songs with non-binary metres are sung by a male chorus.

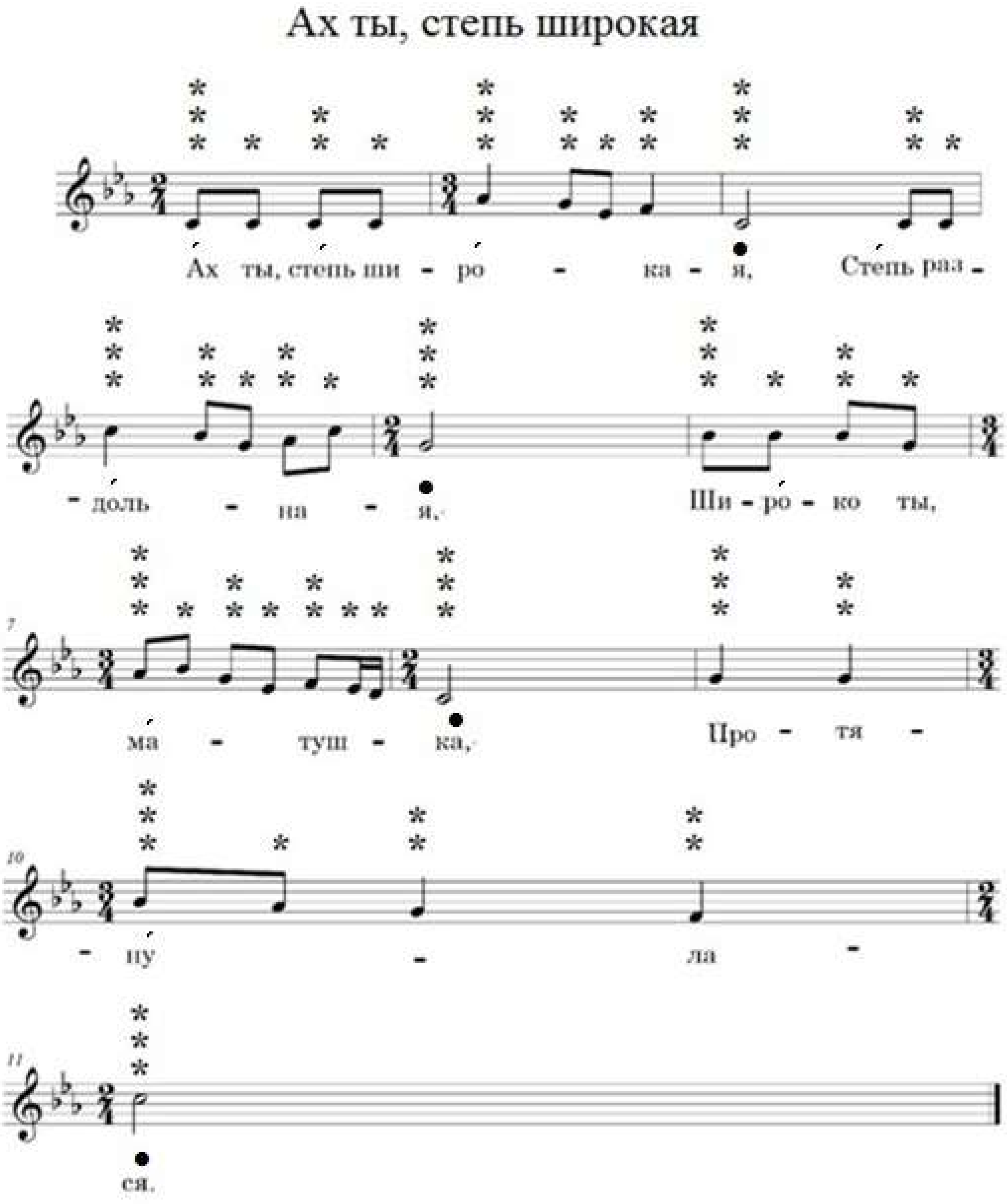



Figure 3.13: A score of an example from ‘Ах ты, степь широкая’ (Oh You, Wide Steppe). The score is notated by Finale 2010. The asterisks show musical prominences. The symbol ● suggests tail rhyme of the lyric.

Figure 3.13 (above) shows that the metre in Russian music can frequently change, although such metric irregularity is not necessarily frequent in Russian music. The text is a Russian verse, where odd lines are seven syllables and even lines are five syllables. Each line of verse is divided into four measures, as a minimal formal unit of music is four measures in most types of music. However, such allotment is not found in the example of the Cossack song shown in Figure 3.13 (e.g. the first note of measure 6 is musically prominent, but the lyric in that position is not a stressed syllable), although the last syllable of a line is expressed by long notes, which means that tail rhythm (expressed by ●) is musically expressed. These long notes suggest that they are end rhymes, whose syllabic duration is longer than other syllables in reading. Looking at the linguistic and musical prominences in Figure 3.13, they do not necessarily co-occur, whereas linguistic prominences of the text in English music (such as Figure 3.3 in Section 3.2) are also musically prominent in song. Additionally, the metre or time signature often changes in Figure 3.13, whereas such frequent change of time signature is exceptional in other European music.

Other examples of Cossack music with non-binary rhythm (metre) are shown in Figures 3.14 and 3.15. In addition to frequent metre change, irregular meters, such as 10/4 and 14/4, are used. In most types of music, the basic metres are 2/4, 3/4, 4/4, or 6/8, and the metre is unchangeable until the end of the piece. The example above was a non-binary metre in the sense that the metre changed frequently, but, as shown in the scores, complicated metres such as 10/4, 12/4, 13/4, 14/4, 15/4, 6/8, 9/8, 12/8, and 18/8 are used, and the metres are changeable within the piece.

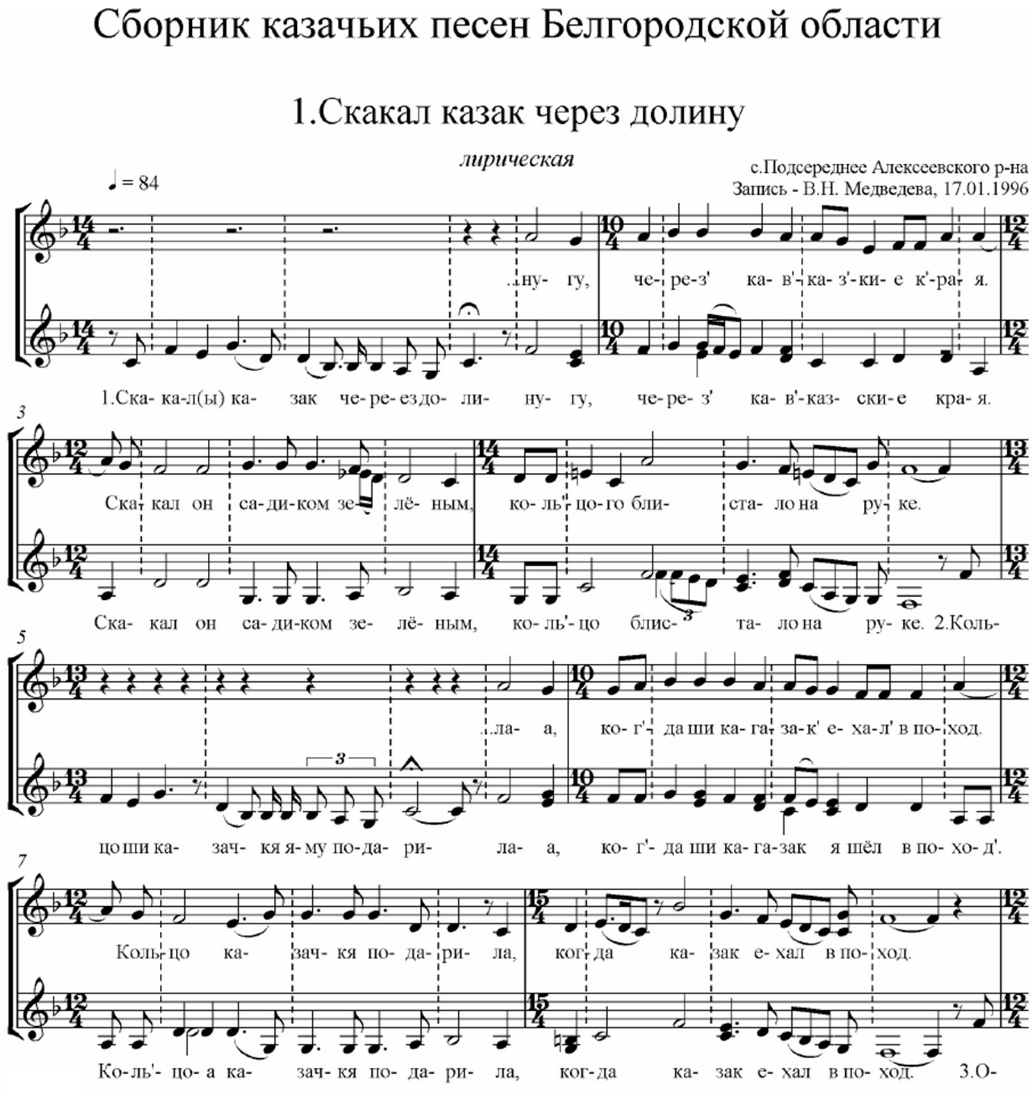



Figure 3.14: A score of an example from ‘Скакал казак через долину’ (The Cossack was crossing the valley). The score is cited from ‘Сборник казачьих песен Белгородской области’ (Collection of Cossack songs of Belgorod region) (2014). A metrical grid is not provided here because the metre of the work is too uneven to present indisputable grid notation.

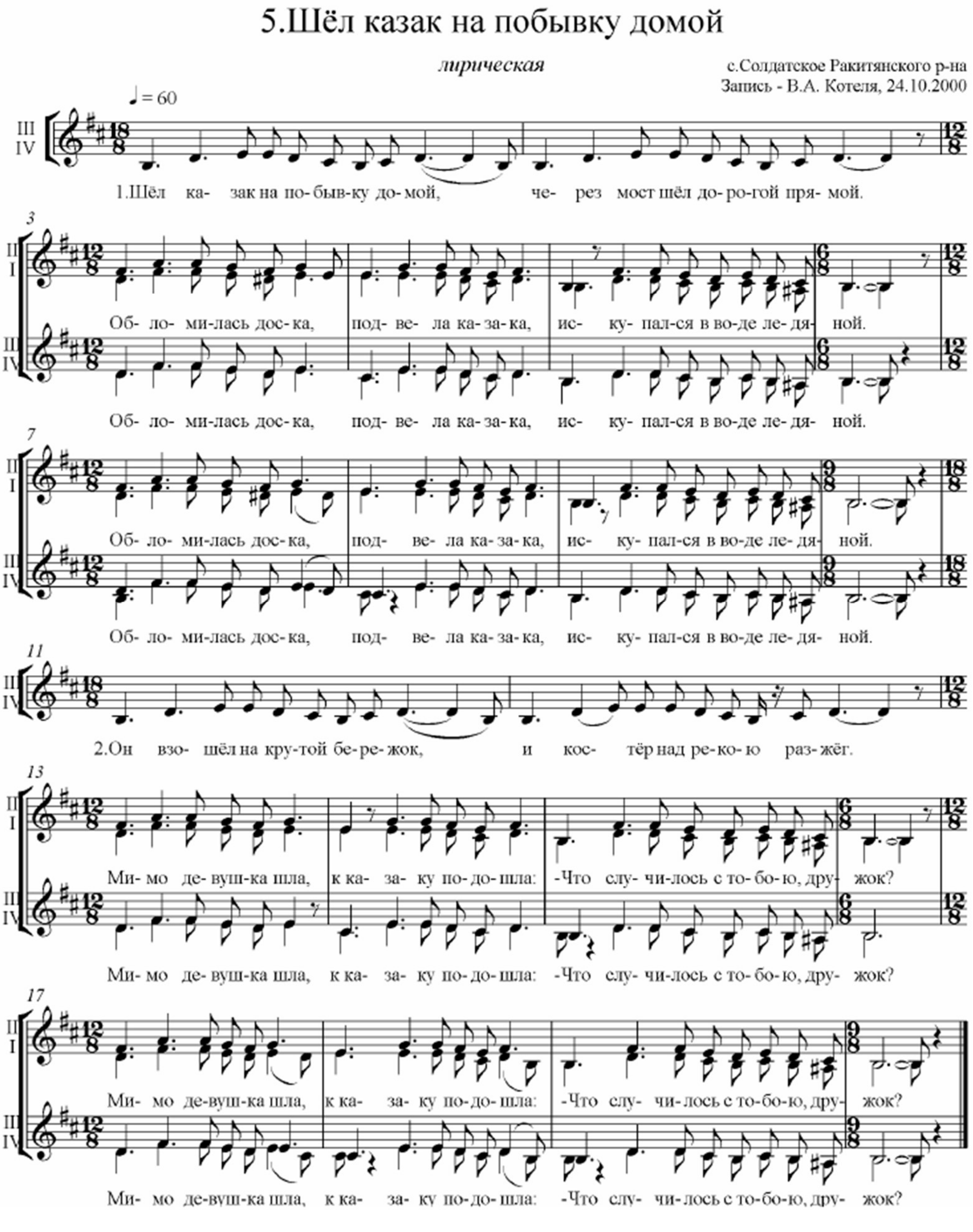



Figure 3.15: A score of an example from 'Шёл казак на побывку домой' (The Cossack was on his way home). The score is cited from 'Сборник казачьих песен Белгородской области' (Collection of Cossack songs of Belgorod region) (2014). A metrical grid is not provided here because the metre of the work is too uneven to present indisputable grid notation.

These types of complicated metre, as shown in Figure 3.15, are generally used in solo vocal work without accompaniments, due to the difficulty of keeping time with other performers or singers. However, the examples above are works for vocal ensemble or choirs, in which singers have to synchronise with each other.

As shown in this chapter, non-binary rhythms or metres were found in some Russian musical works, and such works can be heard on TV or radio programmes in Russian. In other words, Russian speakers are, more or less, familiar with the non-binary rhythms in daily life. Russian musical education also popularises non-binary rhythms. The Russian ministry of education encourages the study of Russian folk music in the national curriculum and, in a similar way to Japan, the Russian government controls the content and hours of music education.

In Russia, authors of textbooks offer a guideline for teachers of the desirable hours of teaching for each subject. This guideline is checked by the Ministry of Education. Teachers in Russia then write a year-long plan of teaching based on the guideline. After that, the headteacher of the school checks the plan to ensure it is based on the guidelines and the national curriculum. Subsequently, all students can equally learn Russian folk music.

The Employee Social Network of Education (Социальная сеть работников образования) and (Копилка уроков) provides examples of a guideline approved by government. According to a guideline for music lessons in the Employee Social Network of Education, seven-year-old pupils study music for a total of 34 hours per year and eight hours of music teaching is for Russian traditional music. The details of the teaching hours are outlined in Table 3.4.

Table 3.4: Teaching hours of music in Russia

Age	Total teaching hours of music	Teaching hours for Russian traditional music
5	72 hours	No description
6	72 hours	No description
7	34 hours	8 hours
8	34 hours	5 hours
9	35 hours	2 hours
10	35 hours	2 hours
11	35 hours	No description
12	35 hours	No description
13	35 hours	No description

Five- and six-year-olds enter infant school, where the curriculum is designed and approved by the ministry, like that of primary and secondary education. In Russia, infant school is not compulsory; however, according to text by RIA Novosti RT (government-funded newspaper), 86% of three- to six-year-olds enrol in infant school.

As the teaching hours and contents of classes, including those in infant schools, are standardised by the ministry and government, all Russian pupils study music in a similar way to the example listed in the table above. However, the detailed curriculum may differ from region to region as the detail of the curriculum can depend on the municipal government. Although the example above is a guideline written by editors and authors of a textbook, actual plans written by teachers might differ from this. The Employee Social Network of Education also offers examples of teachers' plans, these plans were almost identical to the guidelines. For instance, a plan written by N. N. Zhukov (Жукова Н.Н.), 'Education program of music for 7–10-year-old pupils for 2014–2015' (Рабочая программа по музыке в 1- 4 классах на 2014-2015 учебный год), mentions that music teaching hours for seven-year-old pupils were 35 hours, which is similar to the guidelines mentioned above.

It seems that the Russian educational system is similar to the Japanese one, where the ministry of education decides contents and teaching hours for each class. Another similarity between Japan and Russia is that rhythms and metres in Russian and Japanese works of traditional music are complicated, and the metres can change within a piece. Compared with the educational system and rhythms in Japan and Russia, England is different to these countries. First, the programme for music education can be very different from school to school in England, which suggests that some students may not study music to a significant extent. Moreover, rhythms and metres in English music were binary, different to those of Japanese and Russian, where complicated rhythms and metres were used.

One of the differences between Japanese and Russian musical environments was that non-binary rhythms, or complicated metres, were not used in Russian children's songs, while some Japanese works for children had non-binary metres. Except for the children's song, Japanese and Russian musical environments seem to be similar.

3.6 Summary of the musical environments in England, Japan, and Russia

Comparing Sections 3.2, 3.3, and 3.4, it is clear that teaching hours for music are highest in Japan, with the second highest being Russia, and the lowest being the UK. Table 3.5 shows a summary of teaching hours for music in each of these countries.

Table 3.5: Comparison of teaching hours of music between England, Japan, and Russia per year

Age	Teaching hours in England per year	Teaching hours in Japan per year	Teaching hours in Russia per year
5	30 hours	No data (Infant school)	72 hours
6	30 hours	68 hours	72 hours
7	33 hours	70 hours	34 hours
8	33 hours	60 hours	34 hours
9	33 hours	60 hours	35 hours
10	33 hours	50 hours	35 hours
11	36 hours	50 hours	35 hours
12	36 hours	45 hours	35 hours
13	36 hours	35 hours	35 hours
14	Not compulsory	35 hours	35 hours
Total	300 hours	473 hours	422 hours

Although the table above does not include the Japanese teaching hours of music in infant schools, Japanese schools spend many times for music education. Teaching hours of music in the UK are shorter than the other two countries, and teaching hours in reality would be shorter than the data listed in the table above, because academies in the UK do not follow the national curriculum and tend to reduce teaching hours spent on music (Section 3.2).

The data does not include teaching hours for students aged 15 or above. In England, Japan, and Russia, music education for students aged 15 or above is not compulsory; therefore there is no detailed data on the percentage of students in this group who study music in school. However, as it is early music education that mostly affects musical abilities, as shown by Hannon and Trehub (2005) and Prieto (2012), it would be unnecessary to provide such data, and the table above provides the basic tendencies of music education in these countries.

Looking at traditional music in England, Japan, and Russia, Japanese and Russian music both consist of non-binary rhythm, while the rhythm is a constant binary in English music. Japanese and Russian ministries of education emphasise the education of traditional music and offer strict plans for the teaching of it, while the British national curriculum does not provide an exact plan for traditional music, which means that some teachers may not teach traditional music at all.

Based on the summaries above, if musical experience and education affect rhythmic sensitivity, Japanese and Russian people may be sensitive to non-binary rhythm, due to their musical education. Moreover, it might be difficult for English speakers to grasp non-binary rhythm, due to their lack of familiarity with non-binary rhythm.

3.7 Summary of the musical and linguistic rhythmic differences between England, Japan, and Russia

Looking at linguistic rhythm, English and Russian have similar rhythmic structures. Patterns of accented and unaccented syllables are the framework of linguistic rhythm, and the accent of these two countries is that of a stress accent. Japanese accent is a pitch accent, whose prominence is mainly perceived by downward pitch move, as shown in Section 2.5. Moreover, all content word in English and Russian have at least one accent, whereas approximately half of Japanese words have no accent. Due to this lack of accent, lapse, an irregular rhythm caused by successive unaccented syllables, is frequent in Japanese (see Section 2.5).

Although linguistic rhythms in English and Russian are similar, Russian musical rhythm is similar to that of Japanese musical rhythm, while English music rhythm is different to both. In England, rhythms in traditional music are binary, but in Japanese and Russian music, the rhythms can be non-binary.

There is no similarity between Japanese and English musical or linguistic rhythms. However, Russian linguistic rhythm is similar to the English rhythm, but Russian musical rhythm is similar to Japanese musical rhythm. Furthermore, English and Russian musical backgrounds (such as music education) are different. Therefore, if it is linguistic experience that mainly affects rhythmic perception, English and Russian speakers should perceive rhythm similarly. However, if it is musical experience that mainly influences the sensitivity to rhythm, the ability of Japanese and Russian speakers to discern differences in rhythm should be similar.

Chapter 4: Corpus-based study of Japanese rhythm

4.1 Introduction

As mentioned in Section 2.2.2, whether Japanese rhythm is mora-timed or consists of alternations of prominent and non-prominent elements, and what these might be, is a controversial matter. The evidence from Poser (1990), who reported trochaic rhythms in Japanese compounds, onomatopoeia, and nicknames, and Iversen et al. (2008), who found that Japanese listeners showed a preference for trochees in general (see Section 1.3), suggests that Japanese rhythm may be trochaic: a regular pattern of the alternation of prominence and non-prominence.

This point is important to this current thesis, so a more detailed investigation of Japanese rhythm is necessary before we can proceed. Therefore, in this chapter, Japanese rhythm is examined in more detail, using a generous corpus of Japanese speech.

In this study, Japanese rhythm was analysed using the Corpus of Spontaneous Japanese relational database (CSJ RDB) offered by the National Institute for Japanese Language and Linguistics (NINJAL). The CSJ RDB consists of approximately 45 hours of recordings, mainly recordings of conference talks and lectures. Speeches by 1, 417 speakers were recorded between 1999 and 2003, as a national project – Spontaneous Speech: Corpus and Processing Technology.

NINJAL and Maekawa (2003) offer information on prosodic labelling in the corpus (https://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/CSJ_rep.pdf). The corpus was labelled by their programme, Clause Boundary Annotation Programme (CBAP), and trained researchers of the NINJAL team, following a strict protocol for annotation. Of the data, 86% was labelled by the programme, while 14% of data was labelled manually. In the documentation, it is reported that 97–98% of the automatic labelling by the programme was correct.

According to Maekawa (2003), pitch accent is labelled as ‘A’ (accent) and not based on dictionary entries regarding where the accent falls, and which words are accented. Instead, it is based on phonetic evidence from f0. In the corpus, the mora before an immediate F0 drop is considered to be accented and labelled as an accented mora. The degree of the immediate F0 drop had to reach a specific level, which was used as a criterion both for automatic accent detection and manual labelling. Boundaries of each morae, including non-syllabic morae (second half of long vowel [/V:/], mora nasal [/N/], and a geminate obstruent [/G/]), were labelled to examine durations of each morae. The boundaries of the non-syllabic morae were decided based on the discontinuity in spectrogram and waveform (i.e. at the boundaries of the non-syllabic morae, spectrogram, and waveform changes); thus, the changing points were considered to be the beginning and ending of non-syllabic morae. As for the second half of the long vowel, the pitch of the second half is lower than the pitch of the first. This pitch difference was used to decide the boundaries of the first and second parts of the long vowel. With respect to the geminate obstruent, airstream from lung stops resulted in empty space in the spectrogram; the beginning and end of the silences (empty space in the spectrogram) were considered to be boundaries of the geminate obstruent.

Database management software, Navicat, was used to open database (db) files offered by the CSJ RDB. The data were then saved in a csv file and further analysed with R (Team, 2013). Word boundaries were annotated in the corpus. In total, 48, 944 labelled words were included in the corpus with durational, accentual, and positional data. The corpus included only words that were up to 10 morae long. They were distributed as follows: 1-morawords = 3, 946; 2-morawords = 8, 180; 3-morawords = 7, 700; 4-morawords = 7, 965; 5-morawords = 7, 197; 6-morawords = 4, 243; 7-morawords = 3, 478; 8-morawords = 2, 784; 9-morawords = 2, 081; and 10-morawords = 1, 370.

In analysing the duration of mora, durations of non-syllabic morae (second half of long vowel [/V:/], mora nasal [/N/], and geminate obstruent [/G/]) were examined separately, because their durations can be different from syllabic morae (see Section 2.2.2 on non-syllabic morae). Considering the moraic system, non-syllabic morae are counted as single morae.

4.2 Durational effects of accent, non-syllabic morae, and accent position

The duration of each mora was analysed with a linear mixed model (LMM), using the lme4 package in R to examine the effects of: mora type (second half of heavy syllables [/N/, /G/, /Vː/] and syllabic morae); accent position (from first to tenth mora of a word), which was annotated by NINJAL in the corpus; and accentuation (accented or unaccented) (dependent variable = duration; fixed factors = mora type, accent position, accentuation; random factor = word), using the formula in (4.1).

(4.1) Script: m1a = lmer(duration ~ accent_position + stress + mora_type + (1|wordid), data=data).

Table 4.1 is a summary of the statistics calculated from the script above. As can be seen in Table 4.1, the LMM for duration revealed the effects of mora type, accent position, and accentuation.

Table 4.1 Summary of fixed effects

Variable	**df**	**F**	**p**
Mora type	3, 204773	12694.33	<0.0001
Moraic position	1, 193811	154.46	<0.0001
Accentuation	1, 193811	154.46	<0.0001

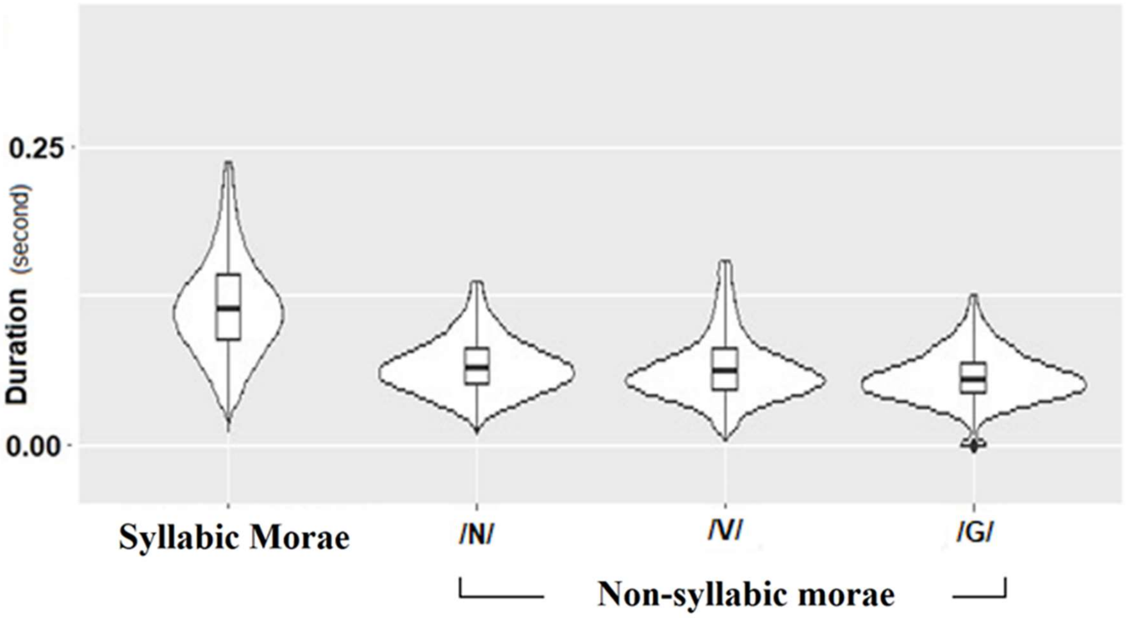


Figure 4.1: A difference of duration between three types of non-syllabic morae (second half of heavy syllables) (/N/, /V:/, and /G/) and syllabic morae.

The violin plots above are a combination of a box plot and a density plot, showing distribution shape, median, and interquartile range. From Figure 4.1 and Table 4.1, it is clear that duration of non-syllabic morae (/N/ [mean = 0.068, SD = 0.03], /V:/ [mean = 0.076, G = 0.06], and /G/ [mean = 0.057, SD = 0.029]) are shorter than syllabic morae (mean = 0.12, SD = 0.0532). Importantly, there is no clear difference between types of non-syllabic morae. Also, looking at the distributional shape of Figure 4.1, durations of non-syllabic morae are less variable (flatter) than normal morae, which means that the duration of non-syllabic morae tends to be constant.

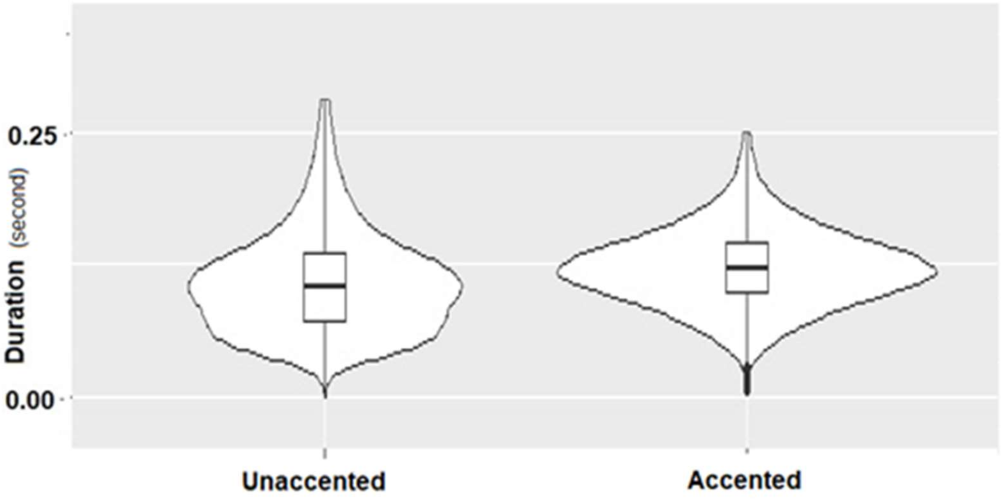


Figure 4.2: A difference of duration between accented and unaccented morae.

Figure 4.2 shows that the duration of accented mora (mean = 0.122, SD = 0.039) is longer than unaccented morae (mean = 0.11, SD = 0.057). Similarly to non-syllabic morae, considering the distributional shape of Figure 4.2, the duration of accented mora is less variable. The analysed data included all types of morae (non-syllabic mora is included here).

Pairwise t-tests were conducted between the moraic positions to examine whether the duration of odd-numbered morae are longer than even-numbered morae, and if durations within the groups (odd-numbered group and even-numbered group) are statistically different (dependent variable = duration; fixed factors = moraic position; random factor = word). The p.adjust = (bonf) function (p-value adjustment by Bonferroni correction) on R was used to calculate adjusted p-values. All pairwise comparisons between durations of adjacent moraic positions were significantly different except for seventh and eighth [mean = 0.111, SD = 0.06, p = n.s.]: first [mean = 0.123, SD = 0.052] and second [mean = 0.103, SD = 0.056][p < 0.0001]; second and third [mean = 0.114, SD = 0.052][p < 0.0001]; third and fourth [mean = 0.102, SD = 0.057][p < 0.0001]; fourth and fifth [mean = 0.113, SD = 0.055][p < 0.0001]; fifth and sixth [mean = 0.107, SD = 0.056][p < 0.0001]; sixth and seventh [mean = 0.112, SD =

0.057][p < 0.0001]; ninth [mean = 0.117, SD = 0.061] and tenth [mean = 0.112, SD = 0.06][p < 0.0036]). The pairwise comparison revealed that there were no statistical differences between second and fourth, third and fifth, fifth and seventh, or eighth and tenth, while there were statistical differences between first and third (p < 0.0001), fourth and sixth (p < 0.0001), and seventh and ninth (p < 0.0001), suggesting that even-numbered moraic durations tend to be similar and odd-numbered moraic durations are not statistically different.

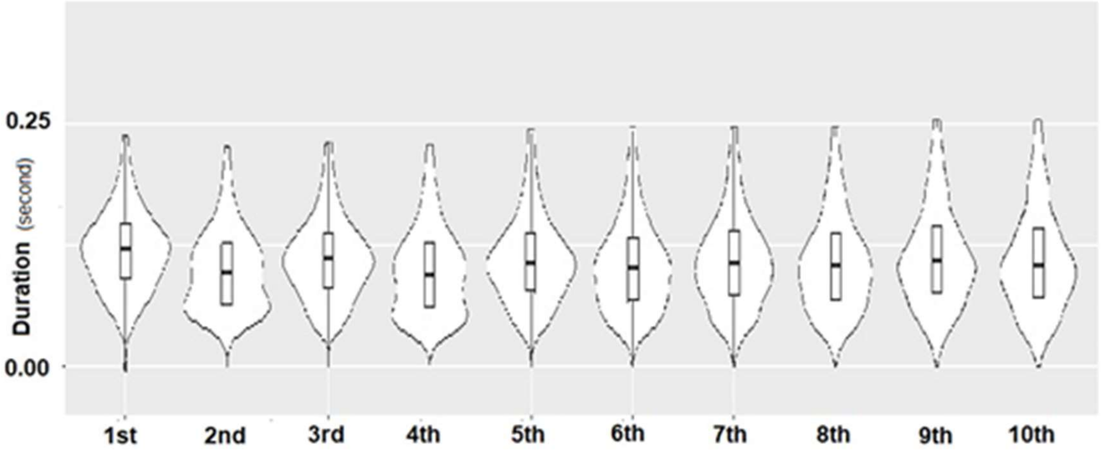


Figure 4.3: A difference of durations between moraic positions.

Figure 4.3 shows mora duration by position in a word. The data is pooled over mora type (non-syllabic morae and syllable morae, accented and unaccented morae). Figure 4.3 and the results of pairwise comparisons seem to show a subtle trochaic rhythm in Japanese.

4.3 Discussion

The analysis showed that accented morae tend to be longer than unaccented morae by, on average, 12 ms (122–110 ms). Furthermore, non-syllabic morae were shorter than syllabic morae by, on average, 48 ms (120–72 ms). Finally, the position of morae affected their duration, in such that odd-numbered morae were longer than even-numbered morae, at least as far as the first four morae are concerned (although the duration of fifth to tenth morae tend to be similar). Also, non-syllabic morae and accented morae tended to be less variable than normal morae and unaccented morae, respectively. These results do not support the idea of mora-timing in Japanese but show that Japanese rhythm is the alternation of prominent and non-prominent morae.

Considering the Japanese trochaic preference shown by Kitahara (2001), Kubozono (2008), Poser (1990), and Rosen (2001), it will be worth examining whether accented morae tend to be on odd-numbered syllables and if non-syllabic morae tend to be in even-numbered positions. To examine this, another analysis will be performed in the following section.

4.4 Position of pitch accent and non-syllabic morae

Similar to the previous analysis, the mora position (from first to tenth morae) was analysed with a logit model in R to examine the relationship between non-syllabic morae, syllabic morae, accented morae, and their position (dependent variable = mora-position; fixed factors = mora type and accentuation; random factor = word), using the script in (4.2). A p-value < 0.05 was considered significant.

(4.2) Script: m1b = glm(mora_position ~ stress + mora_type + (1|tangoid), data=data)

As can be seen in Table 4.2 (below), the model for mora position revealed effects of mora type and accentuation.

Table 4.2 Summary of fixed effects

Variable	**df**	**F**	**p**
Mora position	3, 62328	28.963	<0.0001
Accentuation	1, 59771	45.759	<0.0001

Table 4.3 Percentage of pitch accent and its position

A position of mora	Percentage of pitch accent per moraic position (number of accent/total number of morae in that position)
1st Mora	18.8% (9685/51347)
2nd Mora	9.2% (4381/47421)
3rd Mora	11.1% (4372/39202)
4th Mora	9.6% (3038/31501)
5th Mora	13.7% (3246/23539)
6th Mora	12.8% (1965/16330)
7th Mora	13% (1576/12102)
8th Mora	10.8% (931/8620)
9th Mora	10.9% (638/5832)
10th Mora	10% (379/3754)

In Table 4.3, it is clear that odd-numbered morae are more frequently accented than even-numbered morae. Also, the longer a word becomes, the smaller the difference in the percentage between odd-numbered and even-numbered morae, e.g. the difference between first and second morae is bigger than the difference between third and fourth.

Pairwise t-tests were conducted between types of morae (syllabic and non-syllabic morae) to examine if their positions tend to be different (dependent variable = moraic position; fixed factors = mora type). The p.adjust = (bonf) function (p-value adjustment by Bonferroni correction) on R was used to calculate adjusted p-values. All pairwise comparisons between moraic positions of each morae type were statistically significantly different, except for second half of long vowel and mora nasal (p = n.s.): mora nasal and syllabic morae [$p < 0.0001$]; second half of long vowel and syllabic morae [$p < 0.0001$]; geminate obstruent and syllabic morae [$p < 0.0001$]; geminate obstruent and mora nasal [$p < 0.0001$]; geminate obstruent and second half of long vowel [$p < 0.0001$].

Table 4.4 Percentage of non-syllabic and syllabic morae per position in the words

Position of morae	Percentage of N (mora nasal) per position (number of /N/ / number of total morae)	Percentage of =V: (second half of long vowel) per position (number of /V:/ / number of total morae)	Percentage of G (geminate obstruent) per position (number of /G/ /number of total morae)	Percentage of syllabic morae per position
1st Mora	0.3% (144/51347)	0.02% (10/51347)	0% (5/51347)	99.68% (51188/51347)
2nd Mora	10.7% (5103/47421)	24.9% (11824/47421)	5.1% (2433/47421)	59.3% (28061/47421)
3rd Mora	3.5% (1392/39202)	5.1% (2036/39202)	1.4% (554/39202)	89% (35220/39202)
4th Mora	8.1% (2564/31501)	16.8% (5311/31501)	1.6% (521/31501)	73.5% (23105/31501)
5th Mora	3% (719/23539)	5.6% (1325/23539)	0.9% (216/23539)	90.5% (21279/23539)
6th Mora	6.5% (1072/16330)	10.1% (1652/16330)	1.4% (244/16330)	82% (13362/16330)
7th Mora	3.4% (412/12102)	8.5% (1029/12102)	1% (127/12102)	87% (10534/12102)
8th Mora	5% (436/8620)	12.4% (1077/8620)	0.9% (81/8620)	81.7% (7026/8620)
9th Mora	2.9% (174/5832)	8.6% (505/5832)	1.2% (70/5832)	87.8% (5083/5832)
10th Mora	4.7% (180/3754)	11.1% (417/3754)	1.1% (44/3754)	82.8% (310/3754)

Table 4.4 demonstrates the percentages of all non-syllabic morae. Non-syllabic morae occur more frequently on even-numbered positions compared to odd-numbered positions. As is shown by the pairwise comparisons and Table 4.4, both mora nasal and second half of long vowel tend to be on even-numbered positions, but geminate obstruent are not necessarily on even-numbered positions.

4.5 Discussion

The last result shows that Japanese linguistic rhythm has a clear tendency of trochaic alternative prominence: longer (thus, acoustically more prominent), accented, and syllabic morae are found in odd-numbered positions, whereas shorter (thus, acoustically less prominent), unaccented, and non-syllabic morae are found in even-numbered positions. Only the geminate obstruent shows different tendencies compared to the second half of the long vowel and mora nasal, in the last pairwise comparison.

However, another question arises: whether the duration of unaccented, normal morae (e.g. simple morae) differs depending on the position of mora to make the rhythm trochaic. In other words, it is necessary to examine whether Japanese trochaic rhythm is made just by the position of long morae (accented morae or syllabic morae) and short morae (unaccented morae or non-syllabic morae) or by shortening and lengthening morae.

4.6 Relationship between duration of unaccented syllabic morae and moraic position

The main purpose of the analysis here is to examine whether Japanese trochaic rhythm can be made by shortening and lengthening the duration of unaccented syllabic morae. The method of the analysis was mostly identical to the previous analysis. The duration was analysed with a linear mixed model (LMM), using a lme4 package in R to examine the effects of mora position (dependent variable = duration; fixed factors = mora position; random factor = word). A p-value < 0.05 was considered significant.

To examine whether unaccented morae tend to be longer in odd-numbered positions or shorter in even-numbered positions, another analysis was conducted. From the data used in the last two analyses, data for accented morae and non-syllabic morae was removed to examine if the duration of syllabic morae (unaccented syllabic morae) differs depending on their position in the word.

(4.3) Script: m1b = lmer(duration ~ mora_position + (1|wordid), data=data2)

As can be seen in Table 4.5 (below), the LMM for the moraic duration revealed an effect on mora position.

Table 4.5: Summary of fixed effects

Variable	**df**	**F**	**p**
Mora position	1, 167224	171.44	< 0.0001

Pairwise t-tests were conducted between the moraic positions to examine if the duration of odd-numbered morae is longer than even-numbered morae, and whether durations within the groups (odd-numbered group and even-numbered group) are statistically different. The data here include only the non-accented syllabic morae. The p.adjust = (bonf) function (p-value adjustment by Bonferroni correction) on R was used to calculate adjusted p-values (dependent variable = duration; fixed factors = moraic position). Pairwise comparisons between durations of adjacent moraic positions were significantly different, except for fifth and sixth, sixth and seventh, eighth and ninth, and ninth and tenth, which were not: first and second [$p < 0.0001$]; second and third [$p < 0.0001$]; third and fourth [$p < 0.0001$]; and fourth and fifth [$p < 0.0001$]; and seventh and eighth [$p < 0.0001$]. The pairwise comparisons revealed that there were no statistical differences between second and fourth, third and fifth, fourth and sixth, fifth and seventh, and eighth and tenth, while there were statistical differences between first and third ($p < 0.0001$), sixth and eighth ($p < 0.0001$), and seventh and ninth ($p < 0.0001$).

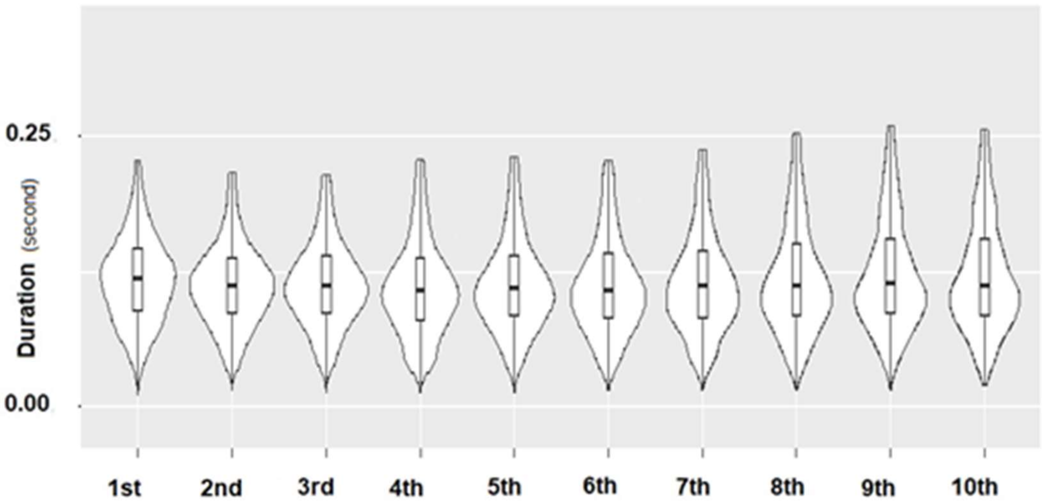


Figure 4.4: A difference of durations between moraic positions (non-syllabic morae and accented morae are excluded).

Figure 4.4 shows the duration of morae in each position, excluding accented morae and non-syllabic morae from data. Looking at Figure 4.4, and the pairwise comparison above, the differences across positions seem to be smaller than in Figure 4.3, which includes accented morae and non-syllabic morae. This suggests that the Japanese subtle trochaic rhythm consists of the long duration of accented morae and the short duration of non-syllabic morae.

It is also important to mention that there is no clear tendency for even-numbered morae to be shorter than odd-numbered morae, which means that the Japanese trochaic rhythm consists purely of the order (positioning) of accented morae and non-syllabic morae.

Table 4.6 Summary of moraic duration (including unaccented syllabic morae)

Position of mora	Average duration and standard deviation (SD)
1st Mora	122 ms (SD [standard error] = 0.054269)
2nd Mora	116 ms (SD = 0.050352)
3rd Mora	118 ms (SD = 0.052326)
4th Mora	116 ms (SD = 0.058752)
5th Mora	118 ms (SD = 0.056795)
6th Mora	117 ms (SD = 0.057549)
7th Mora	120 ms (SD = 0.058636)
8th Mora	123 ms (SD = 0.061228)
9th Mora	126 ms (SD = 0.062276)
10th Mora	124 ms (SD = 0.060497)

Table 4.6 shows that all of these durations are extremely close to each other and all are within the margin of error (they are less than one pitch period and no speaker would be able to keep such small differences consistent).

4.7 Summary of Japanese rhythm based on a corpus

Important findings for the current study are that Japanese rhythm is subtle trochaic and that the trochaic rhythm is achieved by positioning long morae (accented or syllabic morae) in odd-numbered positions and short morae in even-numbered positions, not by adjusting moraic duration for the trochee. This means that the Japanese trochaic rhythm (alternation of prominent and non-prominent morae) is achieved by placing pitch accents mostly in odd-numbered positions and by having non-syllabic morae, which are shorter than syllabic morae, mostly in even-numbered positions.

These results suggest that Japanese rhythm is a subtle trochaic alternation of prominent and non-prominent morae and that the pitch accent works as the foundation for rhythmic structure in a similar way to stress in English.

Although such trochaic rhythm is not always found, due to the fact that not all Japanese content words have an accent, the general Japanese rhythm template might be a simple pattern of prominent and non-prominent morae, similar to that of duple musical metres, which exist in almost all cultures.

Importantly, lapse more frequently occurs in Japanese than English, whose rhythm constrains lapses by adding a stress accent or shifting the accent. Instead, as shown in the results in Figure 4.4, the successive unaccented morae have a mora-timing-like rhythm, and this would explain why it is believed that Japanese is a mora-timed language. This lapse makes Japanese rhythm less binary than English rhythm (i.e. Japanese rhythm does not show even alternations between accented and unaccented mora).

Despite the frequent occurrence of lapse, as shown in Table 4.4, considerable amounts of non-syllabic morae are found on even-numbered morae, which means that short morae (non-syllabic morae) create a trochee-like rhythm, even if accented morae (long morae) do not necessarily follow or precede non-syllabic morae. Short morae (syllabic morae) in even-numbered positions and long morae (the accented morae) in odd-numbered positions make the Japanese rhythm trochaic.

As shown in Table 4.4, geminate obstruent is not necessarily evident in even-numbered positions, which means that non-syllabic morae can also be on odd-numbered morae. In other words, the other two non-syllabic morae can be equally on odd-numbered and even-numbered morae, but they tend to be on even-numbered morae, as if the trochaic rhythm is eurhythmic in Japanese.

Chapter 5: Methods of psychological experiments

5.1 Introduction

This chapter presents, in detail, the experimental methods used for a series of experiments in exploring the perception of rhythm. The basic design of the perception experiments in the current study is based on Hannon and Trehub (2005). The reason that their method was considered suitable for the current study was its ability to measure tolerance to rhythmic violation and rhythmic accuracy for binary and non-binary rhythms. Tolerance and accuracy are necessary to examine the hypotheses in this thesis (for more on the study by Hannon and Trehub [2005] see Section 1.5).

However, because the goal of the current study is to examine the influence of linguistic and musical experience on rhythm perception, the rhythms of the stimuli are different from those of Hannon and Trehub (2005). Also, while Hannon and Trehub (2005) relied on two sessions to examine the effect of exposure to non-binary rhythm, the current study was based on only one session and focused on comparing participants whose musical and linguistic backgrounds were different. A further difference to Hannon and Trehub (2005) is that they used only musical scores; here, three types of stimuli were used: pure-tones, musical stimuli (piano tone), and speech. Finally, another difference from the Hannon and Trehub (2005) study was that the structure-violating stimuli differed from the familiarisation stimulus, not only by inserting elements but also by removing elements to test for possible differences in ratings to stimuli with clash and lapse (linguistic irregular rhythm mentioned in Section 1.2).

The hypotheses tested in the current study are as follows:

Hypothesis 1: All participants, independently of language, will find clashes worse than lapses: they will rate stimuli with clashes as notably different and will be more accurate at detecting differences when stimuli have clashes, compared to those with lapses. This is because of subjective rhythmisation, which means that listeners can anticipate beats. Therefore, they hear beats as regularly occurring even when they are irregular (Grabe & Warren, 1995; Hannon & Trehub, 2005).

Hypothesis 2: Linguistic and musical experience influence the rhythmic perception. This is based on findings by Hannon and Trehub (2005), who showed that people familiar with non-binary rhythm are better at detecting violations to non-binary rhythmic structure compared to people who are unfamiliar with that kind of rhythm. Considering linguistic and musical rhythms discussed in Chapters 2 and 3, Japanese participants will be better at detecting differences in non-binary rhythm because they have experience of this from language and music. Russians will show intermediate ability (musical experience mostly in non-binary rhythms), and English the lowest ability (little familiarity with non-binary rhythms in both language and music).

Hypothesis 3: Looking at Section 2.2.2., lapses frequently occur in Japanese. Considering this, along with Hannon and Trehub's (2005) findings that participants were more sensitive to familiar rhythm than unfamiliar rhythm, Japanese speakers will be more sensitive to lapses than English and Russian speakers. The subjective rhythmisation mentioned in hypothesis 1 may seem to affect this, but Hannon and Trehub (2005) show that rhythmisation affects the results only in the group whose members are not familiar with the rhythm. Therefore, rhythmisation should not affect Japanese participants, who are familiar with lapses.

Hypothesis 4: Linguistic experience will affect how each group treats linguistic stimuli and musical experience will affect how they treat musical stimuli. As musicians are more familiar with musical rhythm than non-musicians, musicians will be more sensitive to musical stimuli. This type of difference between musicians and non-musicians will be lessened in the experiments with tonal and linguistic stimuli. With regard to music, Russian speakers will be similar to Japanese, and English participants very different from both, as Japanese and Russian speakers have similar musical experiences, while the musical experiences of English speakers are different from both Japanese and Russian speakers. Also, the linguistic experiences of English and Russian speakers are similar.

Hypothesis 5: Considering Drake (1993), who demonstrates that musicians performed better in reproducing complicated rhythm than non-musicians, and that there was no significant difference between musicians and non-musicians in reproducing simple rhythms, we can hypothesise that musicians will perceive non-binary rhythm better than non-musicians, while there will be no difference between musicians and non-musicians in perceiving binary rhythm. Musicians will be more accurate and give higher difference ratings than non-musicians, independently of the language.

5. 2 Participants

Participants were divided into two groups: musicians and non-musicians. In the current study, musicians were defined as members of orchestras or choirs, or students who specialise in music at a university (having had formal music education for more than three years). Russian musicians were not recruited as there were no musical conservatories or university where students could major in music, or professional orchestras or choir groups in Izhevsk (the Russian city where the experiments were conducted).

The musical education and activity of all non-musicians was limited to obligatory music education at school. Musical tests – such as a musical aptitude test (Bentley Test), designed to measure musical capability – were not offered to participants to classify them for musicality. An aptitude test would have been ideal but, in general, musical aptitude was expected from the musician group, rather than the non-musician group, as shown in other studies. For example, Drake (1993), who reported differences in the production and perception of rhythm between non-musicians and musicians, found differences between the groups without the need for an aptitude test.

Participants were recruited through email and social networking services (Facebook and VK [a Russian social networking service]) using advertisements detailing the experiment and requesting native speakers from England, Japan, and Russia, both with and without professional musical experience.

Participant data is summarised in Table 5.1: 30 British non-musicians (17F,13M), 18–27 years old (mean = 23.16, SE = 0.34) and 25 British musicians (15F, 10M) 1826 years old (mean = 21.68, SE = 0.4) were recruited in England; 37 Japanese musicians (29F, 8M) 18–33 years old (mean = 24.5, SE = 0.71); 31 Japanese non-musicians (20F, 11M) 18–31 years old (mean = 23.38, SE = 0.61) were also recruited; 33 Russian non-musicians (22F, 11M) 18–24 years old (mean = 22.84, SE = 0.25) were similarly recruited (all Russian participants were recruited in Russia).

British non-musicians were students of the University of Kent. British musicians were students specialising in music at the University of Kent or were members of the university's choir or orchestra. British participants received £15 compensation. In total, 30 British non-musicians and 25 British musicians were recruited from the University of Kent, UK.

Japanese non-musicians were students of Mie University, Osaka Kyoiku University, or Nagoya Isen (college). Japanese musicians and non-musicians were native speakers of standard Japanese or of the

Kansai dialect, suggesting that none were speakers of syllabeme dialects, whose rhythm and prosody are not comparable with standard Japanese or the Kansai dialect; a rhythmic unit in syllabeme is a syllable, which is different from standard Japanese and Kansai dialect, whose rhythmic unit is mora (see Section 2.2.8 for syllabeme rhythm in Japanese). From the Aichi Prefectural University of Fine Arts and Music, located in Aichi, Japan, 26 Japanese musicians were recruited. From Mie University, located in Mie, Japan, 11 Japanese musicians, and nine Japanese non-musicians were recruited. From Osaka Kyoiku University, located in Osaka, Japan, 12 Japanese non-musicians were recruited. From Nagoya Isen (college), located in Aichi, Japan, 10 Japanese non-musicians were recruited. Japanese musicians were students specialising in music at the Aichi Prefectural University of Fine Arts and Music or at Mie University. Japanese participants were paid 1,500 yen in compensation.

Russian participants were students of Izhevsk State Medical Academy. Russian participants were paid 1,000 rubles in compensation. All experiments were held at the institutions listed.

Table 5.1: Demographics of participants

Group name	Mean age	Standard deviation of age	Mean formal music training (years)	Gender
Japanese musician (18 to 33 years old)	24.5	0.71	12.91	29F, 8M
Japanese non-musician (18 to 31 years old)	23.38	0.61	0.12	20F, 11M
Russian non-musician (18 to 24 years old)	22.84	0.25	0.03	22F, 11M
English non-musician (18 to 27 years old)	23.16	0.34	0.1	17F, 13M
English musician (18 to 26 years old)	21.68	0.4	8.36	15F, 10M

Japanese, one British musician, and one Russian non-musician were excluded from the data, due to failure to comply with instructions; they were expected to use an entire scale from 1 (very similar) to 6 (very different) but these participants chose only 1 or 6 throughout the experiment. Russian compulsory education includes five years of English classes, and Japanese compulsory education offers three years of English education. However, all of the participants, including the English speakers, had not travelled abroad for a period longer than three months, so it can be reasonably assumed that they were monolingual, with some exposure to a second language. All recruited participants self-reported that they were not bilingual. All participants were of self-reported, normal hearing.

5.3 Rhythmic structure of the stimuli

Similar to Hannon and Trehub (2005), there were two types of familiarisation stimuli: binary rhythm and non-binary rhythm. Both were followed by one of the seven types of test stimuli: (1) control; (2) structure-preserving; (3) clash1; (4) clash2; (5) lapse1; (6) lapse2; and (7) lapse3 (see Table 5.2). Control test stimuli have an identical rhythm to familiarisation stimuli. Structure-preserving stimuli are made by removing an element but extending another element to twice the length of the other elements to keep the rhythm structure. The rhythmic structures of the stimuli mentioned in this section apply to all stimuli (linguistic, musical, and tonal). This was done to facilitate comparisons across modalities.

The duration of an element (sound) in the stimuli was 200 ms, except for the 400 ms element used in the structure-preserving stimulus (the long structure-preserving element in the linguistic stimulus was a long vowel sound, in the musical stimulus it was a quarter-tone, and in the tonal stimulus it was an additional 200 ms of silence). For example, an entire length of the binary rhythm familiarisation stimulus was 3,200 ms (16 elements × 200 ms).

Table 5.2: List of the rhythmic structure of stimuli (linguistic, musical, and tonal stimuli). ‘$\overset{x}{\text{x}}$’ stands for a prominent element higher or louder than ‘x’, which is a non-prominent element. The symbol ‘:’ suggests that the element preceding this is two times longer than others. This table shows the rhythmic organisation of all sound types (linguistic, musical, and tonal stimuli). The stimuli are available at the following link: http://sumiokobayashi.com/sumiokobayashiresearch.html

Binary Rhythm Familiarisation Stimulus: x x x x x x x x xxxxxxxxxxxxxxxx		
Binary Rhythm Test Stimuli		
Control and SP (Structure-preserving)	Lapse1, 2 and 3	Clash1 and 2
Control: x x x x x x x x xxxxxxxxxxxxxxxx SP: x x x x x x x x x:xxxxxxx:xxxxxx	1: x x x x x x x x xxxxxxxxxxxxxxxxxx 2: x x x x x x x x xxxxxxxxxxxxxxxxxxxx 3: x x x x x x x x xxxxxxxxxxxxxxxxxxxx	1: xx x x xx x x xxxxxxxxxxxxxx 2: xx xx xx xx xxxxxxxxxxxx
Non-binary Rhythm Familiarisation Stimulus: x x x x x x x x xxxxxxxxxxxxxxxxxxxx		
Non-binary Rhythm Test Stimuli		
Control and SP (Structure-preserving)	Lapse1,2 and 3	Clash1 and 2
Control : x x x x x x x x xxxxxxxxxxxxxxxxxxxx SP: x x x x x x x x x:xxxxxxxxx:xxxxxxxx	1: x x x x x x x x xxxxxxxxxxxxxxxxxxxxxx 2: x x x x x x x x xxxxxxxxxxxxxxxxxxxxxxxx 3: x x x x x x x x xxxxxxxxxxxxxxxxxxxxxxxx	1: xx x x xx x x xxxxxxxxxxxxxxxxxx 2: xx xx xx xx xxxxxxxxxxxxxxxx

Table 5.2 suggests all rhythm types of the stimuli. ‘$\overset{x}{\text{x}}$’ is a prominent element whose F0 is higher and/or whose amplitude was louder than ‘x’ (a non-prominent element) depending on sound type (see the details of prominence and stimulus types in Section 5.4).

The binary rhythm familiarisation stimuli had a frequent metric pattern found in English and Russian, and in English music. The non-binary rhythm familiarisation was a non-binary compound metre (2+3/4), which is a possible rhythm in the Japanese language, due to the lack of accent in content words (see Section 2.2.7 for Japanese rhythm), and in Japanese and Russian music. This non-binary rhythm is rarely found in English and in English music.

The test stimuli were variants of familiarisation stimuli, being different from the familiarisation stimulus, due to the addition or deletion of one or two elements: for instance, in Table 5.2, clash1 was made by removing the second element from the familiarisation stimulus.

The position of deletion and addition from familiarisation for clash and lapse were identical both in binary rhythm stimuli and non-binary rhythm stimuli. Jones et al. (1982) suggest that, in judging the pitch differences that are of prominence in musical and linguistic stimuli, the sensitivity to pitch can be different depending on the position in a phrase. Palmer and Krumhansl (1990) also mention that memory of the position of a tone is better when the tone is located in a metrically strong place (i.e. downbeat in a musical phrase). These studies imply that the position of addition and deletion of elements, both in the binary rhythm and non-binary rhythm test stimuli, need to correspond with each other: for example, both in binary and non-binary rhythm test stimuli, second elements are removed from familiarisation stimuli to create clash rhythm, as shown in Table 5.2.

A lapse is frequently found in Japanese speech, as is mentioned in Section 2.2.2. Therefore, lapse test stimuli are suitable to examine whether familiarity with language-specific rhythm affects rhythm perception. Compared with lapse, clash is relatively rare in English, Japanese, and Russian.

A structure-preserving test stimulus is an exceptional stimulus, in the sense that the phrasal duration is not different from its familiarisation stimulus, being removed an element while the one proceeding it is twice as long.

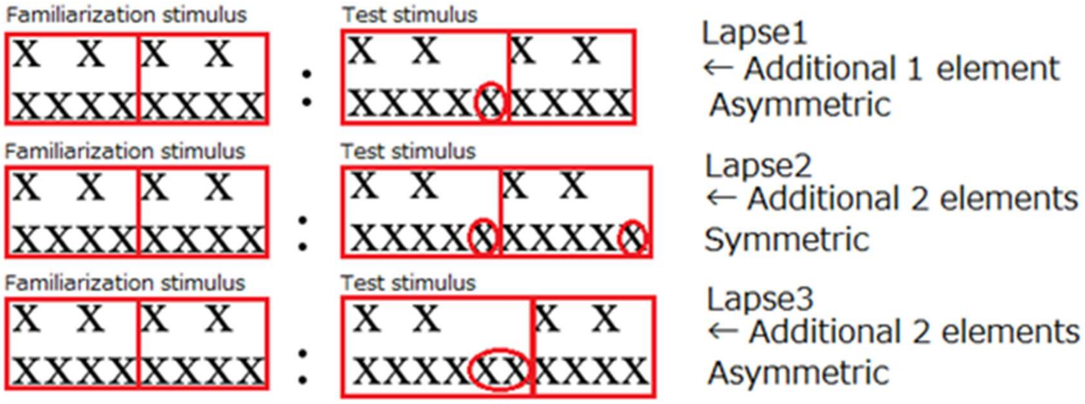


Figure 5.1: Familiarisation and test stimuli (lapse) in binary rhythm trials. The rhythm shown here is the rhythm of all sound types (linguistic, musical, and tonal).

Lapses were of three types (lapse1, lapse2, lapse3). The reason for using three types of lapse was that criteria of rating rhythmic difference may be different depending on culture. As shown by Kitayama, Duffy, Kawamura, and Larsen. (2003), Masuda and Nisbett (2001), McKone et al. (2010), and Nisbett, Peng, Choi, and Norenzayan (2001), some cultural groups perceive a difference depending on whether the stimuli have symmetric structure. Looking at lapse1 and lapse2, lapse2 may seem to be more different from the familiarisation stimulus than lapse1, as it includes two additional elements, whereas lapse1 includes only one. However, some participants from a cultural group may perceive that lapse1 is more different from the familiarisation stimulus than lapse2, as lapse2 and the familiarisation stimulus are both symmetric structures, while lapse1 is an asymmetric structure. This type of difference of viewpoint depends on culture; thus, three types of lapse were used to further examine this hypothesis.

5.4 Stimuli

The acoustic correlates of the prominences in stimulus types are as follows:

1. Linguistic stimuli = fundamental frequency
2. Musical stimuli = fundamental frequency + amplitude
3. Pure-tone stimuli = amplitude

The main purpose of providing different stimulus types was to test whether linguistic and musical experience affect the rhythmic perception of different stimulus types (i.e. to test if musical experience affects only the rhythmic perception of the musical stimulus or all stimulus types, and if linguistic experience affects only the rhythmic perception of linguistic phrase or all stimulus types).

5.4.1 Linguistic stimuli

A female Greek native speaker recorded the /ma/ sounds using a Logitech Webcam C930e. The sampling rate was 44.1kHz. The Greek syllable /ma/ seemed to be optimal for this experiment because [m] is found in all the languages tested here, while the Greek vowel /a/ is phonetically a low central vowel [ɐ] (Arvaniti, 2007) and is different from the low vowel(s) of all the languages tested in the current study. The low central vowel is transcribed as [ɐ] in Greek; therefore, it works well for all three languages tested here, as it is not identical to any of the vowels in English, Japanese or Russian; however, it is not entirely unfamiliar to any speaker group either. This means that no group of speakers would be more familiar with the stimuli than others or would be able to identify them as stimuli from their own language. At the same time, using /ma/ allowed for smooth F0 manipulation.

After recording, the duration of the syllables was adjusted to 200 ms, using a cut function and two select functions in PRAAT (Boersma & Weenink, 2011): the start of selection was moved to nearest zero crossings, and an end of the selection was moved to nearest zero-crossing functions. Prominent (accented) syllables had a high falling pitch (from 252 Hz to 186 Hz), while unaccented [ma] had a flat pitch (186

Hz). An exceptional syllable in the stimuli was a long syllable [maː], which was used for structure-preserving stimuli. The long syllable was also synthesised by PRAAT, using the copy and paste function and the same procedure mentioned previously for the 200 ms syllable. In the manipulation of the lengthening, only /a/ was lengthened, while /m/ was not.

Pitch was synthesised using the manipulation function in PRAAT, with the following settings: time step = 0.05, minimum pitch [Hz] = 75, maximum pitch [Hz] = 600).

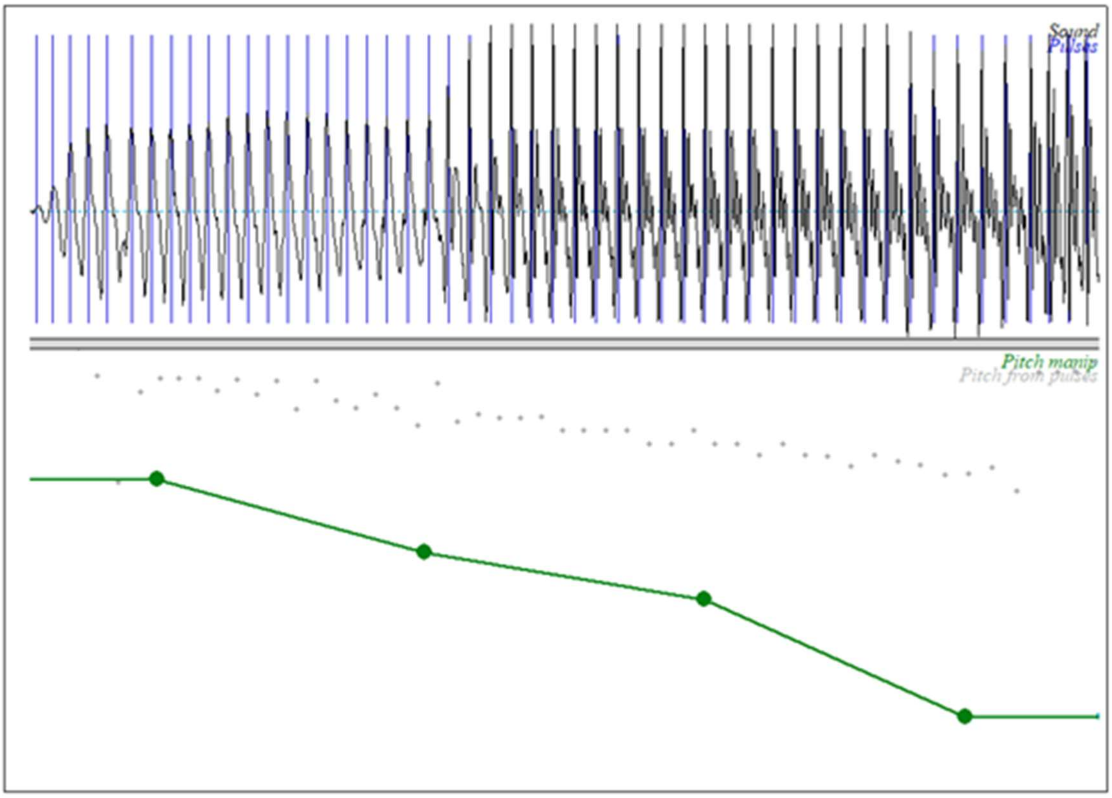


Figure 5.2: F0 manipulation in PRAAT.

In PRAAT, the four green points shown in Figure 5.2 were manually shifted to adjust the F0 of the stimuli using the manipulate function on R. The setting for the manipulation was standard (default). The amplitudes of stressed and unstressed syllables were manipulated to create identical 64 dB using the decibels amplify function in WaveSurfer 1.8.p4 (https://wavesurfer-js.org). These prominent (falling pitch) and non-prominent (flat pitch) sounds were then combined using the concatenate function in PRAAT.

As each syllable was 200 ms, the total duration of stimuli was a multiple of 200 ms. For example, the total duration of the familiarisation stimulus with binary rhythm was 3,200 ms as it consisted of 16 syllables. Figure 5.3 shows the F0 pitch and duration of linguistic stimuli (successive prominences).

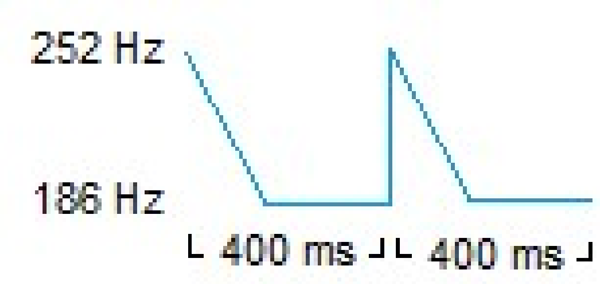


Figure 5.3: F0 movement and duration.

5.4.2 Musical stimuli

The rhythms in all musical stimuli are identical to the rhythms listed in Table 5.2 (Section 5.3). Thus, rhythmically, the musical stimuli had the same structure as the linguistic stimuli. First, the stimuli were notated in Finale 2010, a digital audio workstation and notation software (https://www.finalemusic.com/). Then, the export function in Finale 2010 was used to export the files in wave format. In exporting, the human playback function (which approximates the exported music into human performance) was not used, as it slightly changes the duration and amplitude of each sound and the aim here was to keep the musical stimuli comparable to the linguistic and tonal stimuli, where all elements had standard durations.

In the Finale notation, each sound was written in an eighth note, while the accompaniment that underlines the downbeat, was quarter note or dotted quarter note, corresponding to the distance between downbeats (see Figure 5.4). Figure 5.4 shows the basic rhythmic structure with the accompaniment. As the accent symbol (>) suggests, the left-hand accompaniment underlines metre structure. The stimuli are downloadable here: http://sumiokobayashi.com/sumiokobayashiresearch.html.

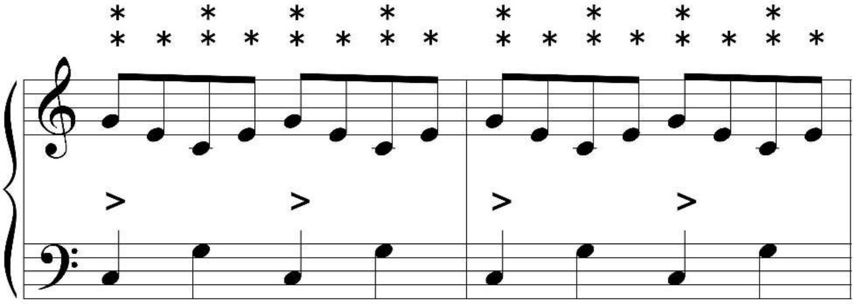


Figure 5.4: Note structure and a metrical grid of the musical familiarisation stimulus with binary rhythm.

The amplitude of the downbeat (prominence) in the musical stimuli was 67 dB. The amplitude of unstressed notes was 64 dB, similar to the amplitude of non-prominent elements in linguistic and tonal stimuli. This is so that participants could listen to different stimulus types in similar conditions. These amplitudes were manipulated with WaveSurfer 1.8.p4, using the decibels amplify function. The duration of each note was, as in the other stimulus types, 200 ms for short notes and 400 ms for long notes (long notes were used in structure-preserving stimuli). An accent symbol (>) in Figures 5.4 and 5.5 indicates that the note is downbeat (accented).

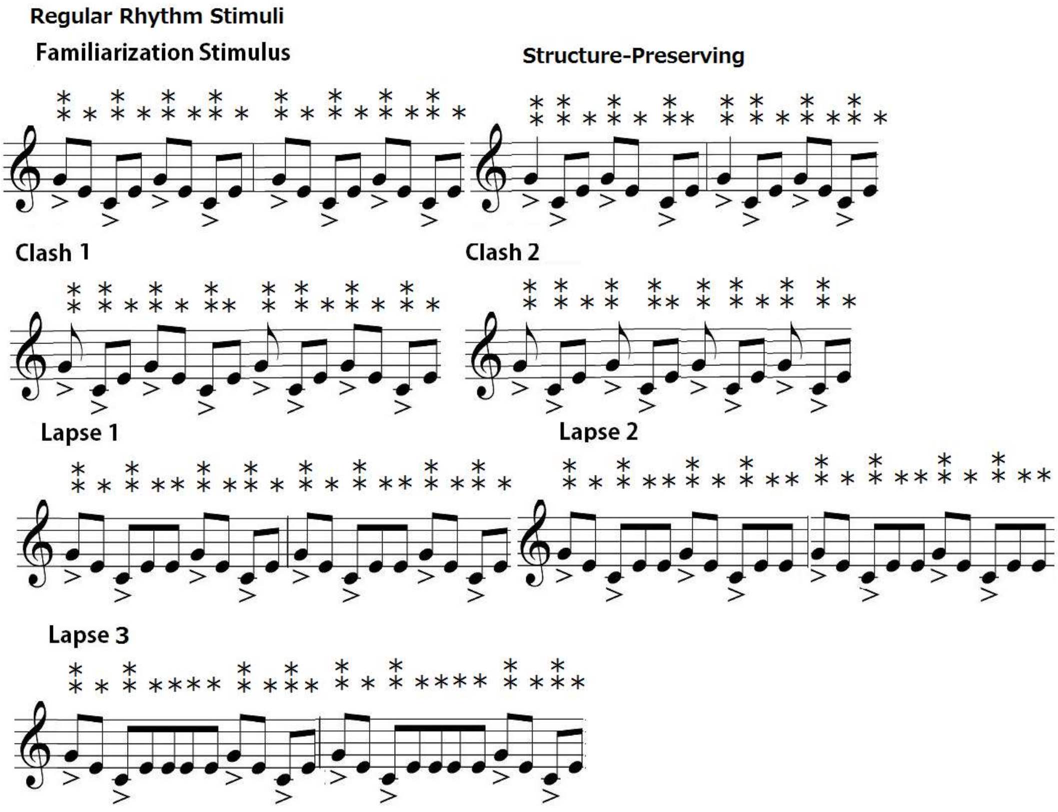


Figure 5.5: Notations and metrical grids of each stimulus (binary rhythm).

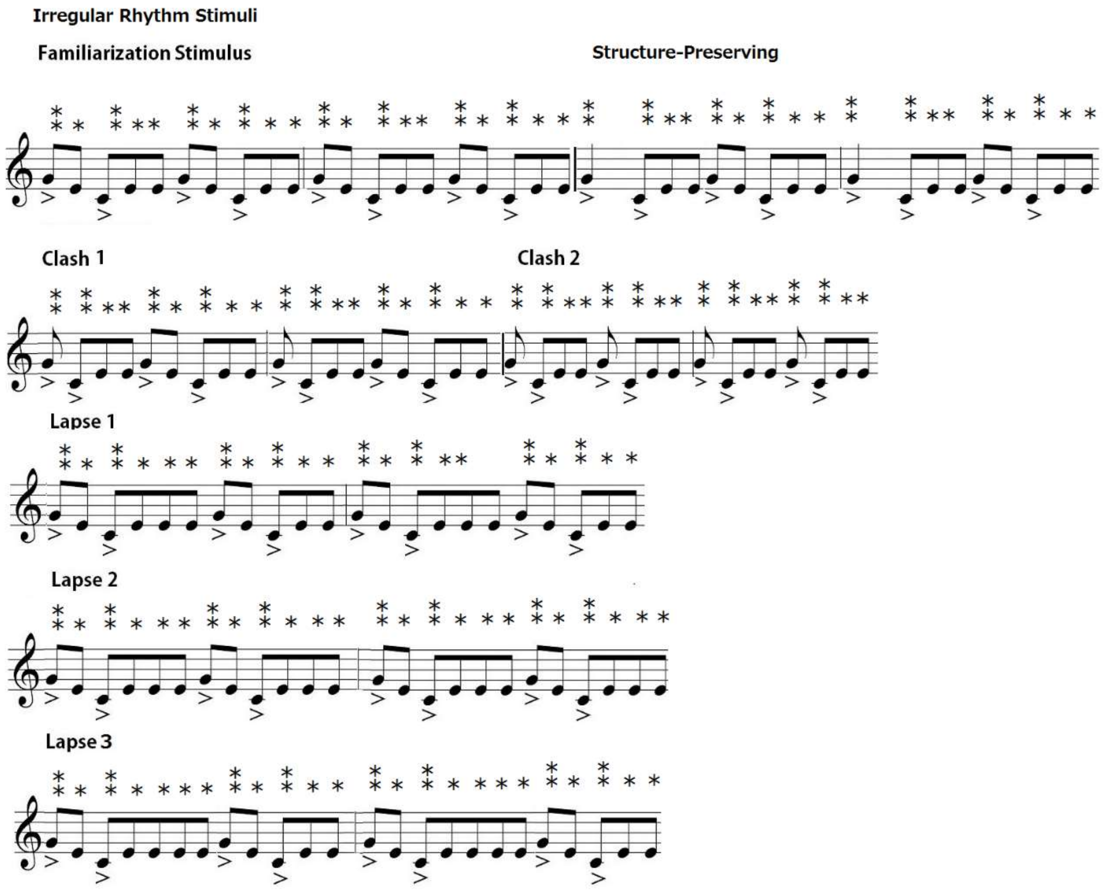



Figure 5.6: Notations and metrical grids of each stimulus (non-binary rhythm).

The harmonic structure or cord did not change throughout the musical phrases, as comprehension of harmony would enhance rhythmic perception. Change of harmonic structure usually suggests the presence of a phrase boundary or rhythmic grouping in music (Bharucha & Krumhansl, 1983). Therefore, a musician's understanding of harmony helps them to grasp the metre.

The musical stimuli were not excessively melodic, as some functional notes, such as leading tone, would be prominent even when they are not stressed (Cuddy, 1989; Dawe et al.1993). Huron and Royal (1996) and Lerdahl and Jackendoff (1983) revealed that big pitch leap (step) can also cause prominence. For this reason, excessive steps were also avoided. Relatively high and low notes are perceived more prominently than others, as is shown in Huron and Royal (1996) and Lerdahl and Jackendoff (1983); therefore, peak and bottom sounds are used only at the downbeats (prominences). Also, the downbeats, indicated by the accent symbol (>), were reinforced by a single note accompaniment (C or G as shown in Figure 5.4). When the melody downbeat was C, the accompaniment was G, and the accented G of the melody was underlined by an accompaniment sound C.

5.4.3 Tonal stimuli (pure tone stimuli)

The rhythmic structures of the tonal stimuli were identical to those of the musical and linguistic stimuli (see Figure 5.2, Section 5.3). The function Create Sound as tonal in PRAAT was used to produce the tonal sounds in the default setting. Short tonal stimuli consisted of 50 ms tone and 150 ms silence, while long tonal stimuli were comprised of 150 ms sound and 50 ms silence. Long sounds in structure-preserving stimuli (400 ms) were expressed by additional long silence (200 ms). At first, two types of tonal sound were needed (50 ms and 150 ms) to create stimuli; these were made by changing the end time setting of the Create Sound as tonal function. The two types of silence (50 ms and 150 ms) were also made using this function, by changing the amplitude value of the setting to 0. The pitch is constant at 440 Hz. The amplitude of accented and unaccented tones was manipulated in WaveSurfer 1.88p4, using the Amplify by Decibels function; accented tones were louder (77 dB) than unaccented tones (64 dB). Pitch and duration were not altered throughout tonal trials.

The length of an element in the tonal stimuli was identical to linguistic and musical stimuli (400 ms for a long element of structure-preserving stimuli and 200 ms for all other elements of the stimuli). Although the lengths of elements are identical regardless of stimulus type, two types of tonal stimuli were used: long tonal stimuli and short tonal stimuli. The difference between short and long tonal stimuli is the lengths of tone and silence that create the borders between the elements. For the long tonal stimuli, each element was 150 ms tone and 50 ms silence, while an element of short tonal stimuli consisted of 50 ms tone and 150 ms silence (see Figures 5.8 and 5.9 for the difference between short and long tones). The element for structure-preserving stimuli was twice as long as the other elements, due to an additional 200 ms silence for both short and long tonal stimuli.

The reason for the difference between short and long tonal stimuli is that rhythmic perception may differ depending on the duration of the sound, as silence is treated as a syllabic unit in Japanese (Hattori, 1980). One of the difficulties faced when designing the tonal stimuli was that the perception of silence inserted between tones to make tonal boundaries might be perceived differently depending on the native language. One of the features of Japanese is a non-syllabic mora, a geminate obstruent, which is counted as a single mora, as argued by Hattori (1980) and discussed in Section 2.2.2. As Kawahara (2015) mentions, disyllabic (bimoraic) words with a geminate, such as [kitte] 'postage stamp' (きって) is counted as three morae, while [kite] 'come' (きて) is counted as two morae. One of the differences between [kite] and [kitte] is that the duration of constriction, or closure duration, of [kitte] is longer than [kite], which means that the closure duration (silence) between syllables in [kitte] is longer. This silence between syllables is counted as a mora in Japanese and changes the meaning of words (Hattori, 1980). To cope with this issue, two types of tonal stimuli (long and short) were prepared.

After producing two types of tone (50 ms tone and 150 ms tone) and silences (50 ms silence and 150 ms silence), the concatenate function in PRAAT was used to produce the stimuli. The 150 ms tonal sound was concatenated with 50 ms silence as a unit of long tonal stimulus, and 50 ms tonal sound with 150 ms silence were combined into a unit of short tonal stimulus. These units were again combined by the concatenate function to create stimuli of strings of tones separated by silence.

Although there were two sorts of tonal stimuli, inter-onset duration remained a constant 200 ms, both in short tonal stimuli and in long tonal stimuli. The exception to this was the 400 ms tone, which was 200 ms longer than other tones, due to an additional 200 ms silence used for structure-preserving stimulus, where an additional 200 ms silent sound file was added to the first tone of the stimulus.

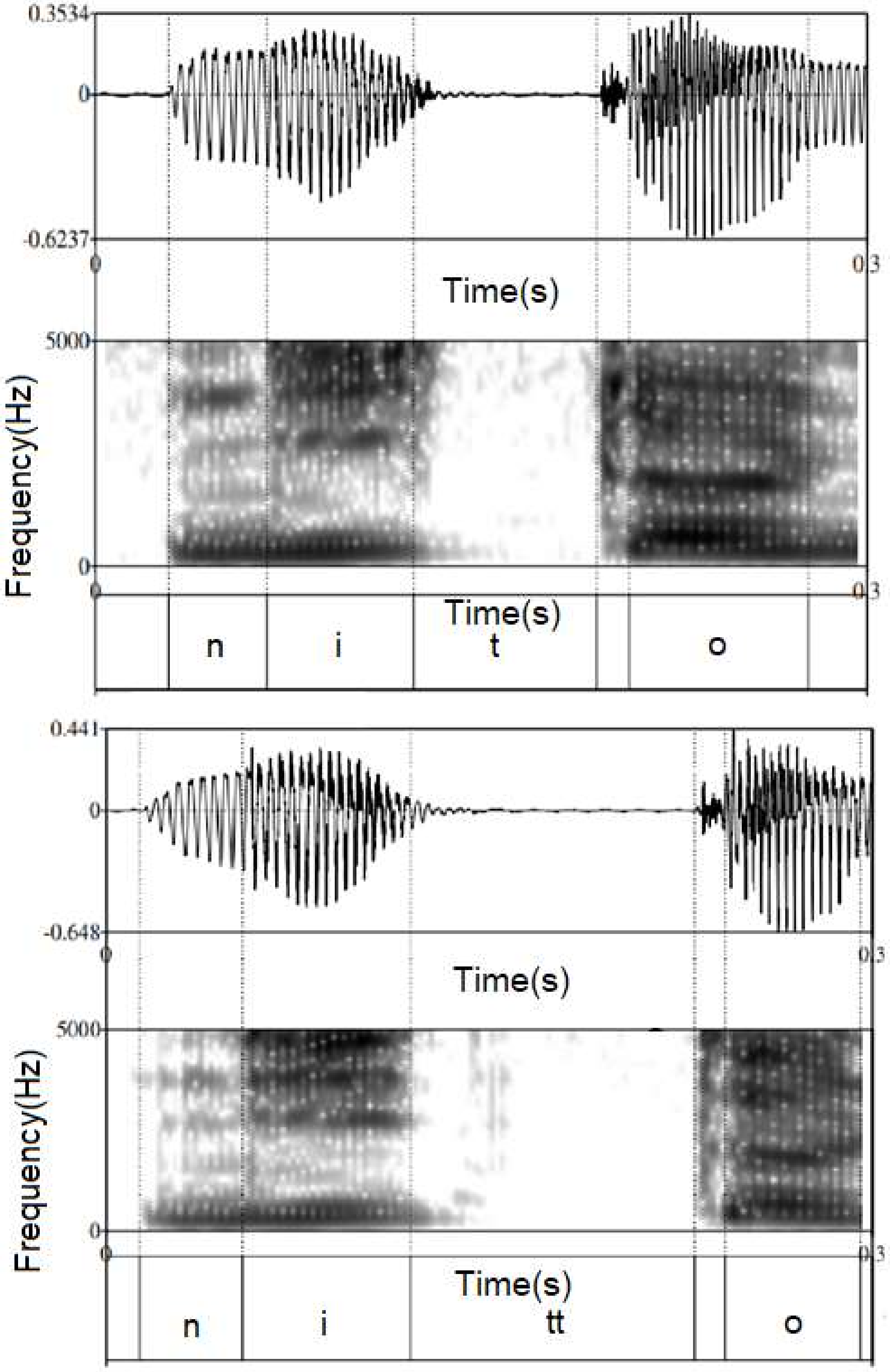


Figure 5.7: A difference between a singleton [t] and a geminate [t] in Japanese from Kawahara (2015).

Figure 5.7 shows the waveforms and spectrograms of a geminate and singleton. Japanese speakers distinguish singleton from geminate based on the duration of the stop (silence) and, as summarised by Kawahara (2015), acoustic correlate and primary cue of the geminate is a duration of silence between the syllables.

Silent intervals are involved in the perception of geminate stops and mora count in Japanese; thus, it was necessary to consider the duration of the silent intervals between tones, as they could be interpreted as being part of the rhythmic structure of the stimuli. Figure 5.8 shows a spectrogram of short tonal stimuli (familiarisation stimulus of binary rhythm).

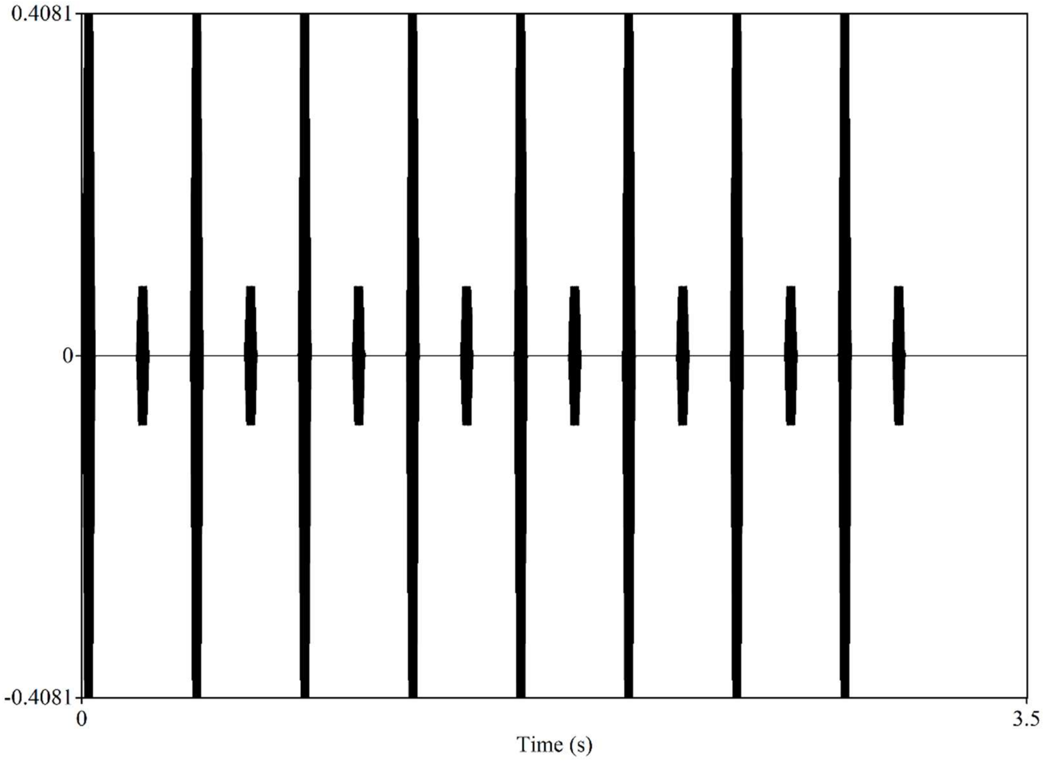


Figure 5.8: Spectrogram of short tonal stimuli.

In order to cope with the difficulty of the possible influence of silence on how Japanese listeners perceived rhythm, the second type of tonal stimuli included shorter silent intervals (50 ms) than those used in the experiment (see Figure 5.9).

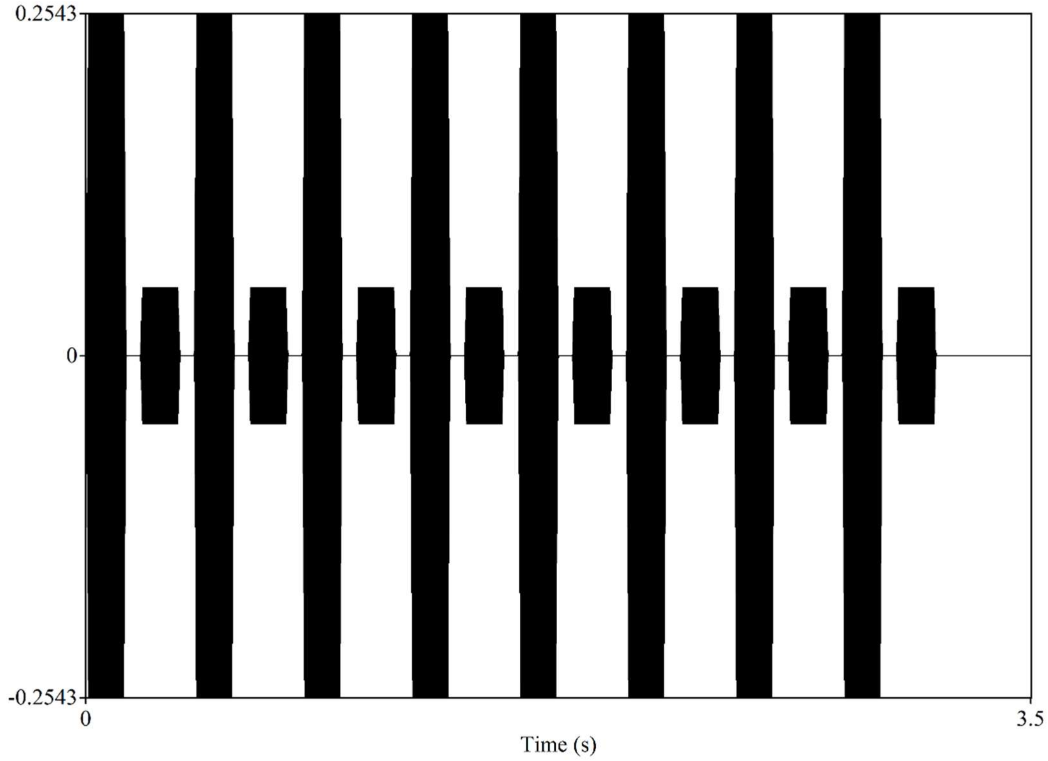


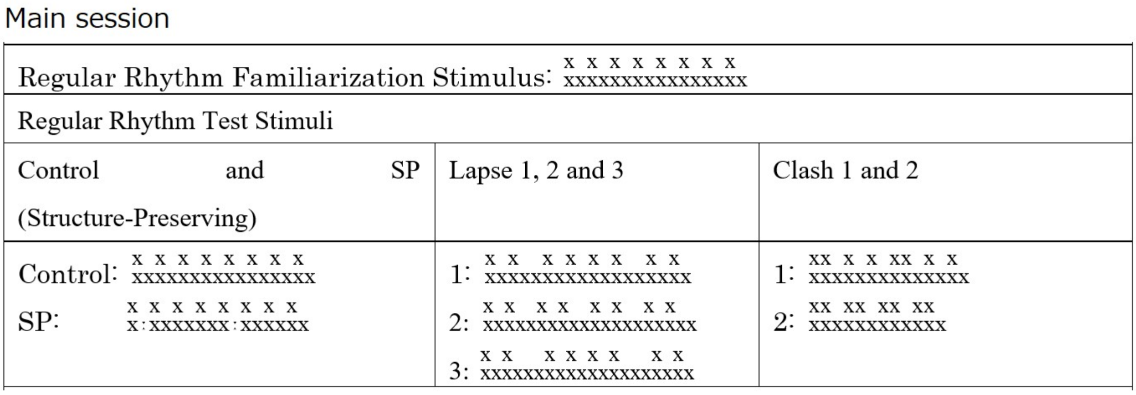
Figure 5.9: Spectrograms of long tonal stimuli.

5.5 Structure of the experiments

Before the main experiment, participants had the chance to practice with 16 practice trials. The stimuli for these trials were comparable but different to the test stimuli (see Table 5.3). The familiarisations used in the practice session were identical to the main session.

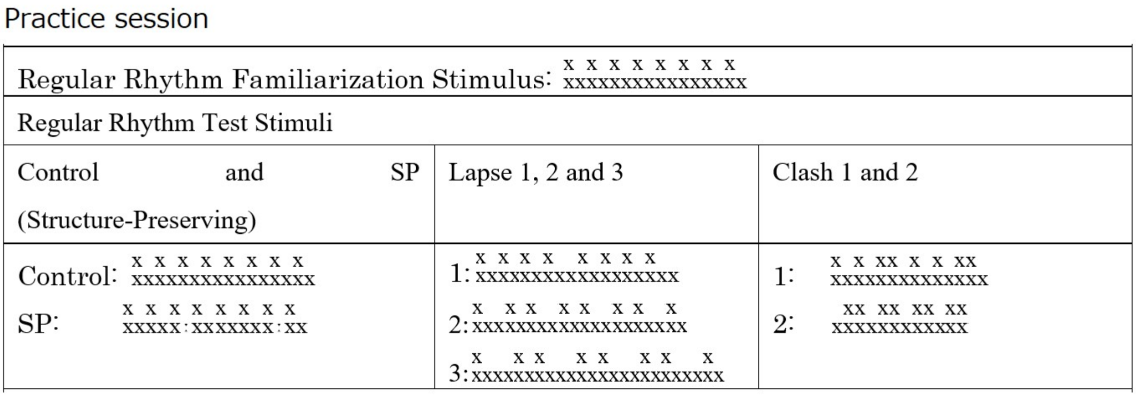
The practice trials consisted of four sessions (linguistic, musical, tonal long, and tonal short stimuli), without a break between them. The linguistic practice trials consisted of binary control, binary lapse1, non-binary control, and non-binary clash1. The musical practice trials consisted of binary control, binary lapse2, non-binary control, and non-binary clash2. The short tonal practice trials consisted of binary control, binary clash1, non-binary control, and non-binary lapse1. The long tonal practice trials consisted of binary control, binary clash2, non-binary control, and non-binary lapse3. In total, the practice experiment consisted of 16 trials, lasting around three minutes. The practice session was held before the main experiments (linguistic, musical, tonal experiment). Participants could take a short pause between the practice session and the main experiment.

Table 5.3: Comparison of rhythms in the main session and practice session.



Main session

Regular Rhythm Familiarization Stimulus: x x x x x x x x xxxxxxxxxxxxxxxx		
Regular Rhythm Test Stimuli		
Control and SP (Structure-Preserving)	Lapse 1, 2 and 3	Clash 1 and 2
Control: x x x x x x x x xxxxxxxxxxxxxxxx SP: x x x x x x x x x:xxxxxxx:xxxxxx	1: x x x x x x x x xxxxxxxxxxxxxxxxxx 2: x x x x x x x x xxxxxxxxxxxxxxxxxxxx 3: x x x x x x x x xxxxxxxxxxxxxxxxxxxx	1: xx x x xx x x xxxxxxxxxxxxxx 2: xx xx xx xx xxxxxxxxxxxx



Practice session

Regular Rhythm Familiarization Stimulus: x x x x x x x x xxxxxxxxxxxxxxxx		
Regular Rhythm Test Stimuli		
Control and SP (Structure-Preserving)	Lapse 1, 2 and 3	Clash 1 and 2
Control: x x x x x x x x xxxxxxxxxxxxxxxx SP: x x x x x x x x xxxxx:xxxxxxx:xx	1: x x x x x x x x xxxxxxxxxxxxxxxxxx 2: x x x x x x x x xxxxxxxxxxxxxxxxxxxx 3: x x x x x x x x xxxxxxxxxxxxxxxxxxxxxxxx	1: x x xx x x xx xxxxxxxxxxxxxx 2: xx xx xx xx xxxxxxxxxxxx

The practice and main sessions were run on OpenSesame (version 3.0.0a19) using a notebook, ASUS TransBook T100TA, at a comfortable listening level determined by the experimenter.

Stimuli were presented online through JVC headphones (JVC AXIV HA-S200) to participants seated in a separate sound-attenuated room, where a maximum two participants could attend the experiment. The participants chose an answer by left clicking with a mouse (Logicool m100r) on buttons that were shown on a computer display. When two participants attended the session, two sets of the same equipment were provided for the participants to rate the rhythm difference in the same environment. Similarly, all settings, such as brightness and volume, remained constant throughout all the experiments to test perception under identical conditions.

5.6 Procedure

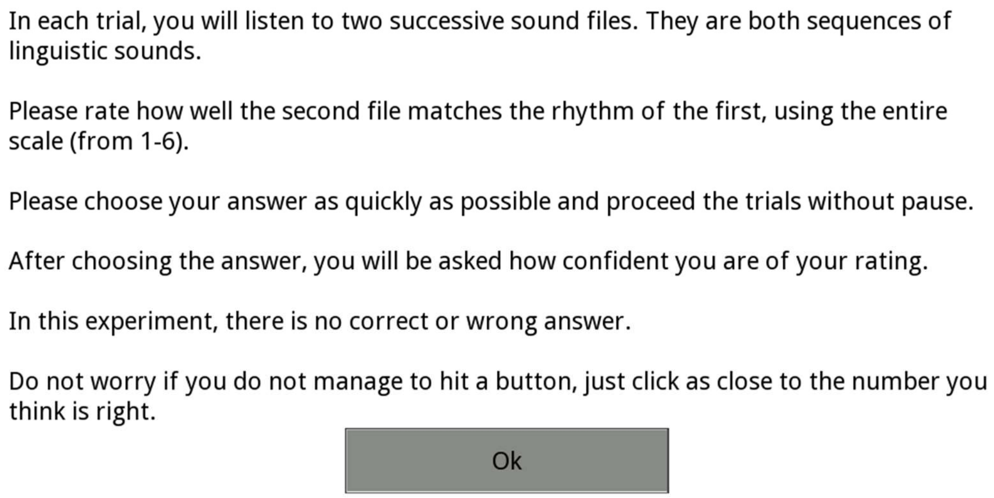
5.6.1 Pilot experiment

Prior to the main experiments, a pilot was held to consider possible problems with the design and set-up. Five Japanese speakers (four females and one male, aged between 18–22 years old) and five English speakers (five males, aged between 20–23 years old) who did not take part in the main experiments were tested at the University of Kent. They were non-musicians. The pilot was identical to the main experiment in terms of set-up. The only difference was the rhythms used in the stimuli: the familiarisation stimuli were identical to the main experiment, but there was no lapse 3.

After finishing the experiment, participants were asked if the procedure was difficult to understand. As there was no negative feedback, the design was kept the same for the main experiment. However, after the pilot, the additional explanation ‘use an entire scale from 1 to 6’ was included to avoid a situation in which participants chose only 1 (very similar) or 6 (very different).

5.6.2 Procedure of the main experiment

Prior to commencing data collection, the participants read an information sheet in a sound-treated room at each university (University of Kent, Mie University, Osaka Kyoiku University, Nagoya Isen, and Izhevsk State Medical Academy) where the experiments were held. The information sheets, written in English, Japanese, and Russian, described the purpose of the experiment and the procedure, and provided contact information for participants to ask questions or withdraw from participating, emphasising that they could refuse to take part, could withdraw at any time, and that their personal information was not kept once they were remunerated. Participants filled in these questionnaires before the experiment and refer the reader to the participants section. The information sheet explained that the participants would hear two sound files divided by 1000 ms of silence, and that they would be asked to judge a rhythmic difference within the pair on a scale of 1 (very similar) to 6 (very different). The phrase ‘rhythmic difference’ was used in the information sheet instead of ‘metre’ or ‘metric difference’ because, while musicians would correctly understand the notion of metre, non-musicians might not understand the meaning of metre and the difference between rhythm and metre. The instructions were also displayed on the monitor at the beginning of each experiment, including the practice experiment. Other instructions displayed were: (1) Please choose your answer as quickly as possible and proceed the trials without pause; (2) After choosing the answer, you will be asked how confident you are of your rating; and (3) The two successive sound files are both sequences of beep sound (or linguistic or musical sound, depending on the session). After reading the instruction, participants clicked OK on the screen using a mouse. Then, the message, ‘Click “start” to continue the experiment’, and a start button appeared (see figures for the actual procedures in Opensesame). Every time a participant clicked on a displayed button, the following screen was displayed. Figures 4.10, 11, 12, and 13 show the actual procedure.


In each trial, you will listen to two successive sound files. They are both sequences of linguistic sounds.

Please rate how well the second file matches the rhythm of the first, using the entire scale (from 1-6).

Please choose your answer as quickly as possible and proceed the trials without pause.

After choosing the answer, you will be asked how confident you are of your rating.

In this experiment, there is no correct or wrong answer.

Do not worry if you do not manage to hit a button, just click as close to the number you think is right.

Figure 5.10: Screen shown in Opensesame at the start of a session.

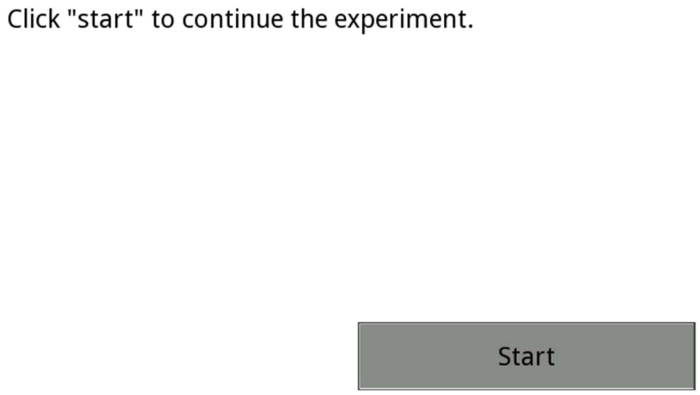
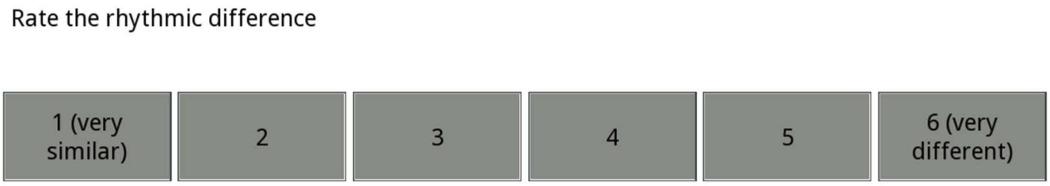
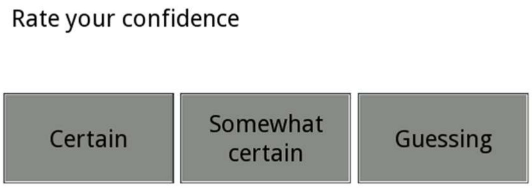

Click "start" to continue the experiment.

Figure 5.11: 'Start' button to start playing stimuli.



Rate the rhythmic difference

Figure 5.12: Buttons to rate rhythmic difference were automatically shown after playing a pair of stimuli (familiarisation stimulus and test stimulus).



Rate your confidence

Figure 5.13: Confidence ratings were shown after rating the rhythmic difference.

Once a participant selected ‘start’, Opensesame automatically played the familiarisation stimulus first, followed by 1000 ms of silence, before playing the test stimulus. Opensesame automatically displayed the rating scale for the difference, displaying the message, ‘Rate the rhythmic difference’. On rating, Opensesame displayed the message, ‘Rate your confidence’ (Figure 5.13). After selection, participants were shown the message, ‘Click “start” to continue the experiment’. This process of clicking ‘start’ was self-paced.

In the main experiments, the cycle was repeated 42 times (7 violation types [control, structure-preserving, clash1, clash2, lapse1, lapse2, lapse3] × 2 rhythm types [binary and non-binary types] × 3 repetitions). There were 42 trials for each of the four stimulus types (linguistic, musical, tonal long, and tonal short stimuli), which meant that the main session consisted of 168 trials.

Following the practice experiment, the experimenter checked that the participant was ready for the main session, then the main session started. There were four sessions, one for each of the four stimulus types, i.e. linguistic, musical, long tonal, and short tonal stimuli. Each session consisted of three blocks (three times repetition) in which 14 trials (14 pairs of sound files) were heard. The order of the 14 pairs of stimuli in a block was randomised. The order of the four main sessions (musical, linguistic, short tonal, and long tonal sessions) was counterbalanced between participants. A session lasted approximately 10 minutes. The practice and four sessions were held within a day from 10 a.m. to 5 p.m., depending on the schedules of the participants. Once a participant finished an experiment, they could take a short break (10 minutes maximum), this was including after the practice experiment. The experimenter asked participants to start a new experiment (different stimulus type [linguistic, musical, or tonal session]) after the short break. Each participant gave 168 ratings (14 pairs × 4 stimulus types × 3 blocks [three repetitions]).

5.7 Measurements and Statistics

All ratings on a scale of 1 to 6 (1 = very similar, 6 = very different) and confidence ratings (certain, somewhat certain, guessing) were recorded in an Excel file, through OpenSesame. Some levels such as clash1 clash2, lapse1, lapse2, lapse3 were collapsed into two levels (clash and lapse) through ‘Sum-coding’ of R, using the ‘contrasts’ function. This was done to avoid overcomplicating the analysis.

5.7.1 Statistics for the linguistic experience comparison group

The participants whose data was analysed were non-musicians, excluding the data of musicians from English and Japanese speakers.

Following Hannon and Trehub (2005), the data were used to calculate accuracy. Specifically, accuracy was calculated by subtracting the mean rating for control stimuli (identical rhythm to familiarisation stimulus) from the mean rating for each structure-violating stimulus (clash1 and 2, lapse1, 2, and 3). This accuracy measurement and the ratings provided by participants on the 1–6 scale were treated as dependent variables in statistical analyses.

The ratings from participants were standardised (z-scoresd) through Microsoft Excel, by subtracting the mean of the given set of data (one participant's ratings) from an individual raw value to be standardised, and then dividing the difference by the standard deviation of the given set of data. The standardised ratings, i.e. the z-scoress (standardised ratings), were further analysed. High z-scores suggest that a participant judged the difference between the familiarisation stimulus and the test stimulus to be significant. Z-scoring was done for each participant.

The z-scoresd ratings and accuracy were analysed with linear mixed models (LMM), using lme4 package in R to examine the effects of violation type (lapse, clash, structure-preserving, or control), native language (English, Japanese, or Russian), rhythm type (binary or non-binary), and musical experience (musician or non-musician) (dependent variables = rating and accuracy; fixed factors = violation type, native language, stimulus type, rhythm type; random factor = participant). LMMs were used because they are robust when testing imbalanced datasets (Harrison et al. 2018), as is the case here: there is an imbalance in the present datasets because the clash condition contains two subsets (clash1 and 2) while the lapse condition contains three subsets (lapse1, 2, and 3). A p-value < 0.05, shown by the best-fit model, with step () function on R, was considered statistically significant. The three types of asterisk (*, **, and ***) used in the figures refer to '$p < 0.05$', '$p < 0.01$', '$p <0.001$', respectively. Post hoc tests, using Tukey's honestly significant difference [HSD] post hoc test, were completed using the 'emmeans' function in R.

The ratings and accuracy of 30 British non-musicians (17F,13M), 18–27 years old (mean = 23.16, SE [standard error] = 0.34) 31, Japanese non-musicians (20F, 11M) 18–31 years old (mean = 23.38, SE = 0.61), 33 Russian non-musicians (22F, 11M) 18–24 years old (mean = 22.84, SE = 0.25) were analysed.

A 3 × 4 × 2 × 7 mixed-factorial design crossed three native languages (English, Japanese, and Russian) with three stimulus types (language, music, and tonal), two rhythm types (binary and non-binary), and four violation types (control, structure-preserving, clash, and lapse). Another design crossed two native languages (English, Japanese, Russian), four stimulus types (language, music, short tonal, and long tonal), two timings (binary and non-binary), and four rhythm types (control, structure-preserving, clash, and lapse).

5.7.2 Statistics for the musical experience comparison group

The musical experience comparison group focuses on whether, and how, musical experience affects rhythm perception by comparing musicians with non-musicians. The data of English musicians, Japanese musicians, English non-musicians, and Japanese non-musicians was used for the statistical analysis in the musical experience comparison. The data of the Russian-speaking participants was excluded, as they did not include a musician group. Statistical methods are similar to those used to test linguistic experience. The main difference from linguistic experience comparison is that musical experience (musician or non-musician) is added as a fixed factor, in addition to violation type, native language, stimulus type, and rhythm type.

The z-scoresd ratings for accuracy were analysed with LMM, using lme4 package in R to examine the effects of violation type (lapse, clash, structure-preserving, or control), native languages (English or Japanese), rhythm type (binary or non-binary), and musical experience (musician or non-musician) (dependent variable = rating and accuracy; fixed factors = violation type, native languages, stimulus type, rhythm type, musical experience; random factor = participant).

A p-value < 0.05 was considered significant. The three types of asterisk (*, **, and ***) used in the figures refer to $p < 0.05$, $p < 0.01$, $p <0.001$, respectively.

Chapter 6: The role of linguistic experience: English, Japanese, and Russian

6.1 Introduction

This chapter focuses on whether, and how, linguistic experiences affect rhythm perception, comparing English, Japanese, and Russian speakers, excluding the data of musicians. The following hypotheses were tested by experiments:

Hypothesis 1: All participants, independently of language, will find clashes worse than lapses: they will rate stimuli with clashes as notably different and will be more accurate at detecting differences when stimuli have clashes, compared to those with lapses.

Hypothesis 2: Considering the familiarity with different types of linguistic and musical rhythm discussed in Chapters 2 and 3, Japanese participants will be best at detecting differences in non-binary rhythm, because they have experience of this from language and music. English will have the lowest ability, because they have limited familiarity with non-binary rhythms in both language and music. Finally, Russians will show intermediate ability and fall between the English and Japanese groups, because they have experience with musical non-binary rhythms, but not in language.

Hypothesis 3: It is clear that lapses occur more frequently in Japanese than English and Russian (see Section 2.4). Considering this along with Hannon and Trehub's (2005) findings that participants were more sensitive to familiar rhythm than unfamiliar rhythm, Japanese speakers will be more sensitive to lapses than English and Russian speakers.

6.2 Statistical Method

The accuracy of the linguistic stimuli was statistically examined through R (Dependent variable = accuracy, Random factor = participant, Fixed factors = native language [English, Japanese, Russian], rhythm type [binary or non-binary], and violation type [clash, lapse, SP]). In the script below, two interactions between the native language and rhythm type and between the native language and violation type were included to examine the hypotheses mentioned in Section 6.1.

(6.1) Script: mb1 = lmer(accuracy ~ native_language*binary_or_non-binary + native_language*violation_type (sp, clash, and lapse) + (1|participant), data=accuracy_linguisticstimuli_languageexperiencecomparisongroup).

Z-scoresd ratings of the linguistic stimuli were statistically examined through R (Dependent variable = rating, Random factor = participant, Fixed factors = native language [English, Japanese, Russian], rhythm type [binary or non-binary], and violation type [clash, lapse, and SP). The script (6.2) below was used for these statistical analyses to examine ratings of linguistic stimuli.

(6.2) Script: mb2 = lmer(zaverage ~ native_language*binary_or_non-binary + native_language*violation_type (SP, clash, and lapse) + (1|participant), data=rating_linguisticstimuli_linguisticexperiencecomparisongroup)

Accuracy of the musical stimuli was statistically examined through R (Dependent variable= accuracy, Random factor = participant, Fixed factors = native language [English, Japanese, Russian], rhythm type [binary or non-binary], and violation type [clash, lapse, and SP]). The script (6.3) below was used for these statistical analyses to examine the accuracies of musical stimuli.

(6.3) Script: ma1 = lmer(accuracy ~ native_language*binary_or_non-binary + native_language*violation_type (SP, clash, and lapse) + (1|participant), data=accuracy_musicalstimuli_linguisticstimulicomparisongroup)

Rating of the musical stimuli was statistically examined through R (Dependent variable= z-scoresd rating, Random factor = participant, Fixed factors = native language [English, Japanese, Russian], rhythm type [regular or non-binary], and violation type [clash, lapse, SP]). The script (6.4) below was used for the statistical analyses to examine ratings of musical stimuli.

(6.4) Script: ma2 = lmer(zaverage ~ native_language*binary_or_non-binary + native_language*violation_type (SP, clash, and lapse) + (1|participant), data=rating_musicalstimuli_languageexperiencecomparisongroup)

The accuracy of the tonal stimuli was statistically examined through R (Dependent variable= accuracy, Random factor = participant, Fixed factors = native language [English, Japanese, Russian], rhythm type [binary or non-binary], and violation type [clash, lapse, SP]). The script (6.5) below was used for these statistical analyses to examine the accuracies of tonal stimuli.

(6.5) Script: mc1 = lmer(accuracy ~ native_language*binary_or_non-binary + native_language*violation_type (SP, clash, and lapse) + (1|participant), data=accuracy_puretonestimuli_languageexperiencecomparisongroup)

The rating of the tonal stimuli was statistically examined through R **(**Dependent variable= z-scoresd rating, Random factor = participant, Fixed factors = native language [English, Japanese, Russian], rhythm type [binary or non-binary], and violation type [clash, lapse, SP]). The script (6.6) below was used for these statistical analyses to examine ratings of tonal stimuli.

(6.6) Script: mc2 = lmer(zaverage ~ native_language*binary_or_non-binary + native_language*violation_type (SP, clash, and lapse) + (1|participant), data=rating_puretonestimuli_languageexperiencecomparisongroup)

6.3 Linguistic Stimuli

6.3.1 Accuracy

As can be seen in Table 6.1, the LMM for accuracy revealed the effects of language, rhythm type, and violation type, and interaction between language and violation type. However, there was no interaction among language and violation types.

Table 6.1 Summary of fixed effects

Variable	**df**	**F**	**p**
Language	**2, 92**	**8.35**	**0.0005**
Rhythm type	**1, 848**	**203.48**	**< 0.0001**
Violation type	**2, 848**	**63.5**	**< 0.0001**
Native_language*Rhythm type	**2, 848**	**8.04**	**0.0003**
Native_language*Violation type	4, 840	1.24	n.s.

Table 6.2: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (calculated by subtracting mean of a right variable from a left variable)**	**df**	**t**	**p**
Clash vs. Lapse	**0.7**	**848. 0**	**4.7**	**<0.0001**
Clash vs. SP	**-0.876**	**848. 0**	**-4.87**	**<0.0001**
Lapse vs. SP	**-1.626**	**848. 0**	**-8.9**	**<0.0001**
English binary vs. Japanese binary	**-0.3**	**136. 0**	**-1.98**	**0.05**
English binary vs. Russian binary	0	136.0	-0.18	n.s.
Japanese binary vs. Russian binary	0.3	136.0	1.77	n.s.
English non-binary vs. Japanese non-binary	**-0.9**	**136.0**	**-5.33**	**<0.0001**
English non-binary vs. Russian non-binary	**-0.4**	**136.0**	**-2.32**	**0.022**
Japanese non-binary vs. Russian non-binary	**0.5**	**136.0**	**2.96**	**0.004**

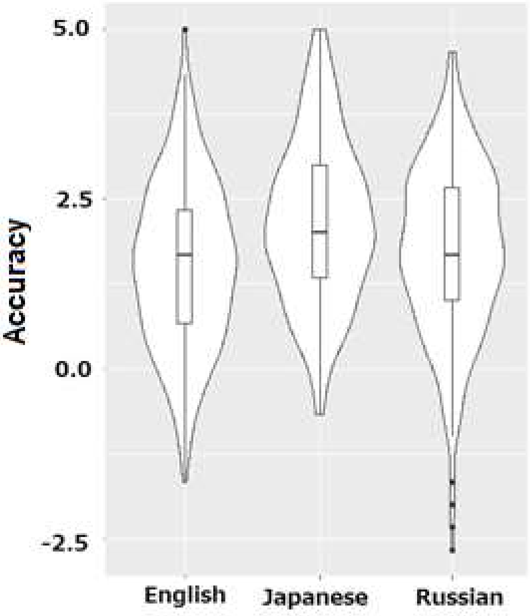


Figure 6.1: Violin plots (showing distribution shape, median, and interquartile range) of accuracy, shown separately for language type. Data from linguistic stimuli only. Data from musicians was excluded.

From Figure 6.1, we can see that Japanese speakers were more accurate than English participants (mean = 2.152, SE = 0.1 vs. mean = 1.536, SE = 0.1, respectively). Similarly, the Japanese group was more accurate than the Russian group (mean = 1.74, SE = 0.1). There was no difference in accuracy between Russian and English participants.

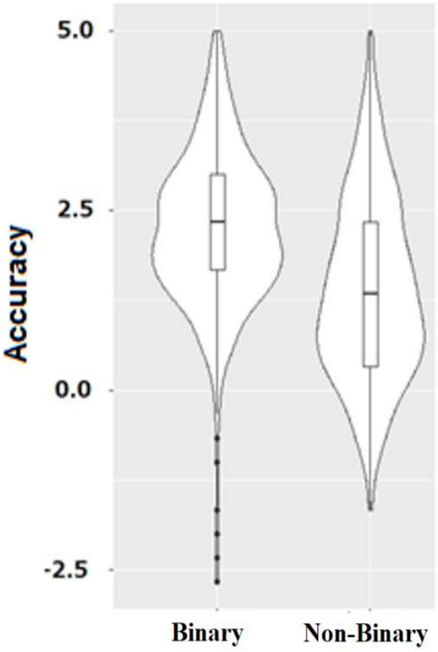


Figure 6.2: Violin plots of accuracy, shown separately for rhythm type. Data from linguistic stimuli only. Data from musicians was excluded.

From Figure 6.2 above, we can see that accuracy was greater for binary (mean = 2.23, SE = 0.069) than non-binary rhythm (mean = 1.393, SE = 0.069).

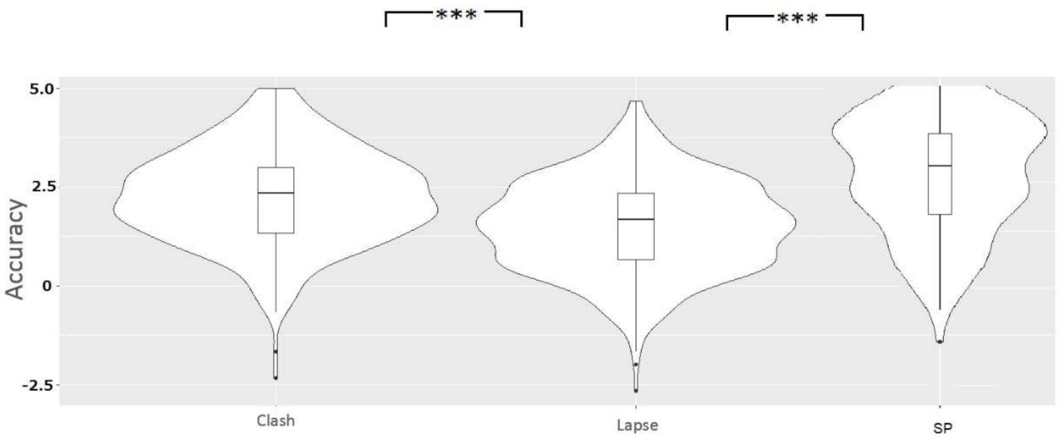


Figure 6.3: Violin plots of accuracy, shown separately for violation type. Data from linguistic stimuli only. Data from musicians was excluded.

From Figure 6.3 and Table 6.2, we can see that rhythm violations due to a clash were more accurately perceived than violations due to a lapse (mean = 2.25, SE = 0.062 vs. mean = 1.5, SE = 0.062, respectively). Accuracy was highest in SP stimuli (mean = 3.126, SE = 0.088). Although it was hypothesised that Japanese speakers would be sensitive to lapse, due to the frequent presence of lapse in their language, there was no interaction between language and violation type.

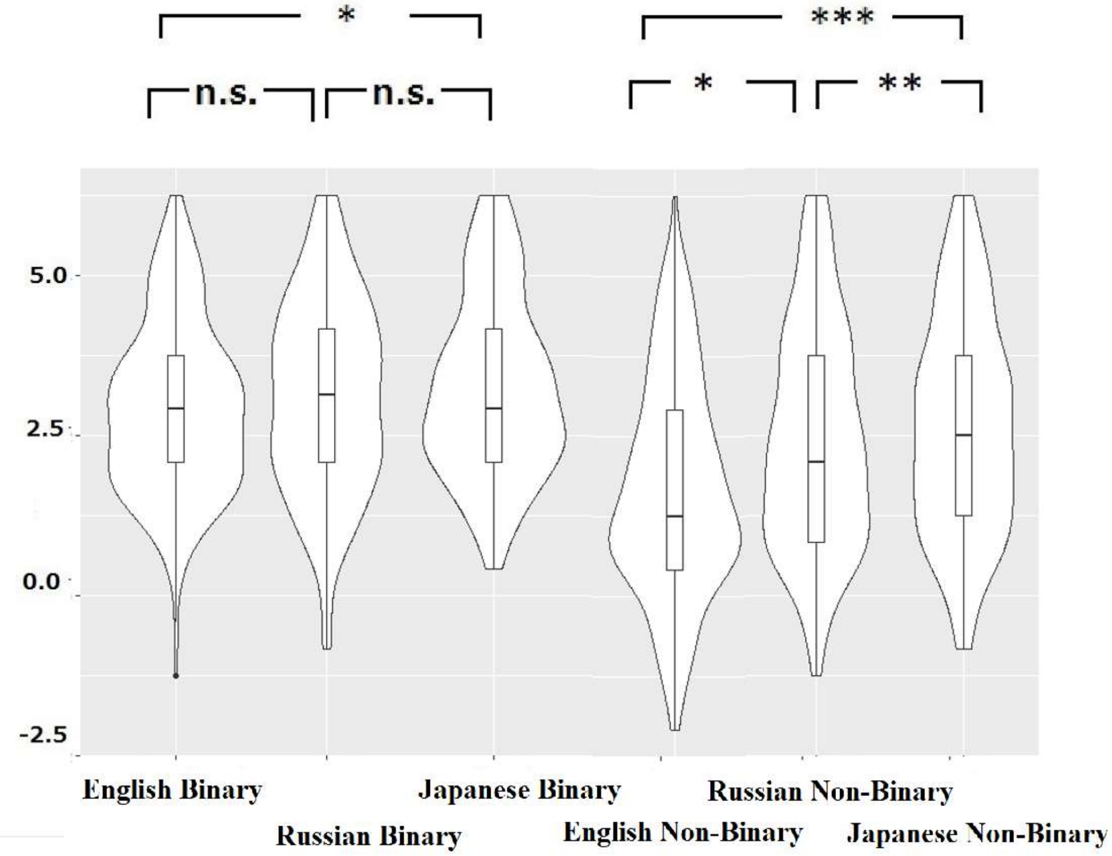


Figure 6.4: Violin plots of accuracy, shown separately for interaction between rhythm type and language. Data from linguistic stimuli only. Data from musicians was excluded.

As shown in Figure 6.4 and Table 6.2, considering accuracy results with respect to binary rhythm, there was no significant difference between English (mean = 2.1, SE = 0.074) and Russian speakers (mean = 2.13, SE = 0.1); however, the accuracy of binary rhythm of the Japanese speakers (mean = 2.44, SE = 0.082) was higher than that of English and Russian speakers. For the accuracies of non-binary rhythm, Japanese accuracy (mean = 1.86, SE = 0.103) was highest, and the second highest was Russian (mean = 1.35, SE = 0.092). English accuracy of non-binary rhythm (mean = 0.96, SE = 0.093) was significantly lower than for the other two groups. The difference between Japanese and Russian accuracies was bigger than the difference between English and Russian accuracies. Although the Russian accuracy of binary rhythm fell between Japanese and English speakers, the difference between the groups was bigger in the non-binary rhythm.

6.3.2 Rating

Table 6.3 provides the summary statistics for the z-scoresd rating of linguistic stimuli. The LMM for the rating revealed the main effects of language, rhythm type, and violation type, interactions between language and rhythm type, and interactions between language and violation type.

Table 6.3 Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Language	**2, 83.67**	**7.93**	**0.0007**
Rhythm type	**1, 819.21**	**45.51**	**< 0.0001**
Violation type	**2, 817.11**	**100.06**	**< 0.0001**
Native_language*Rhythm type	**2, 818.52**	**3.7**	**0.0250**
Native_language*Violation type	**4, 816.65**	**2.41**	**0.0079**

Table 6.4: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
English binary vs. Japanese binary	**-0.3**	**136. 4**	**-2.33**	**0.021**
English binary vs. Russian binary	-0.1	118. 7	-0.72	n.s.
English binary vs. English non-binary	**0.5**	**815. 7**	**5.94**	**< 0.0001**
Japanese binary vs. Russian binary	0.2	126. 9	1.66	n.s.
Japanese binary vs. Japanese non-binary	0.1	829. 0	1.66	n.s.
Russian binary vs. Russian non-binary	**0.3**	**808. 8**	**4.41**	**< 0.0001**
English non-binary vs. Japanese non-binary	**-0.7**	**129. 7**	**-4.8**	**< 0.0001**
English non-binary vs. Russian non-binary	**-0.2**	**118. 0**	**-1.82**	**0.071**
Japanese non-binary vs. Russian non-binary	**0.4**	**123. 2**	**3.05**	**0.003**
English clash vs. Japanese clash	**-0.8**	**379. 4**	**-4.24**	**< 0.0001**
English clash vs. Russian clash	-0.2	331. 2	-0.95	n.s.
English clash vs. English lapse	**0.5**	**810. 9**	**3.58**	**< 0.0001**
English clash vs. English sp	**-1.3**	**816.4**	**-9.26**	**< 0.0001**
Japanese clash vs. Russian clash	**0.6**	**349. 3**	**3.41**	**< 0.0001**
Japanese clash vs. Japanese lapse	**0.6**	**816. 5**	**3.33**	**< 0.0001**
Japanese clash vs. Japanese sp	**-0.6**	**828. 5**	**-3.91**	**< 0.0001**
Russian clash vs. Russian lapse	**0.1**	**809. 3**	**0.47**	**< 0.0001**
Russian clash vs. Russian sp	**-1.1**	**808. 9**	**-8.35**	**< 0.0001**
English lapse vs. Japanese lapse	**-0.6**	**366. 2**	**-3.72**	**< 0.0001**

English lapse vs. Russian lapse	0.3	815. 2	1.87	n.s.
English sp vs. Japanese sp	-0.1	337. 0	-0.73	n.s.
English sp vs. Russian sp	0	290. 6	0.06	n.s.
Japanese sp vs. Russian sp	0.1	324. 5	0.78	n.s.

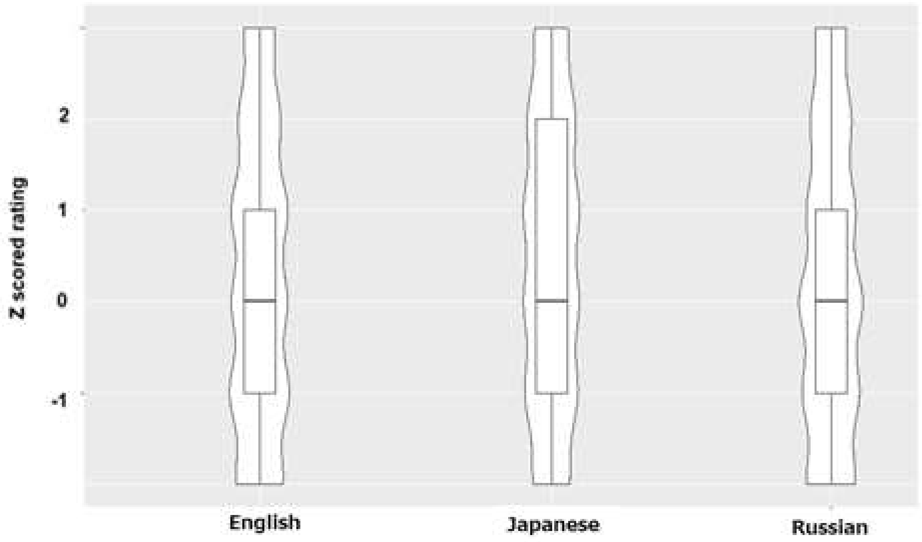


Figure 6.5: Violin plots of z-scores, shown separately for each language. Data from linguistic stimuli only. Data from musicians was excluded.

From Figure 6.5, we can see that the rating for Japanese speakers (mean = 0.805, SE = 0.091) was higher than that for English (mean = 0.317, SE = 0.083) and Russian (mean = 0.483, SE = 0.083) speakers. Similar to the result in the previous section, English and Russian speakers, whose rhythm has similar regularity, had similar ratings.

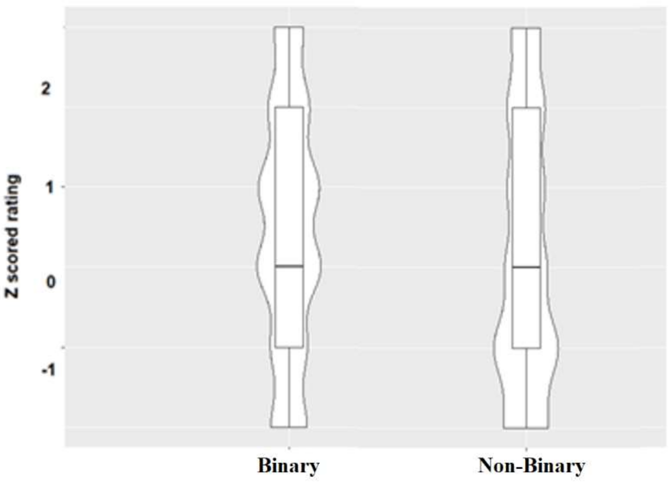


Figure 6.6: Violin plots of z-scores, shown separately for each language. Data from linguistic stimuli only. Data from musicians was excluded.

Figure 6.6 shows that the ratings of violations regarding binary rhythm (mean = 0.695, SE = 0.055) were higher than ratings pertaining to non-binary rhythm (mean = 0.375, SE = 0.054). This suggests that, when participants detected violations in stimuli with binary rhythm, they rated the differences more strongly than differences in stimuli with non-binary rhythm.

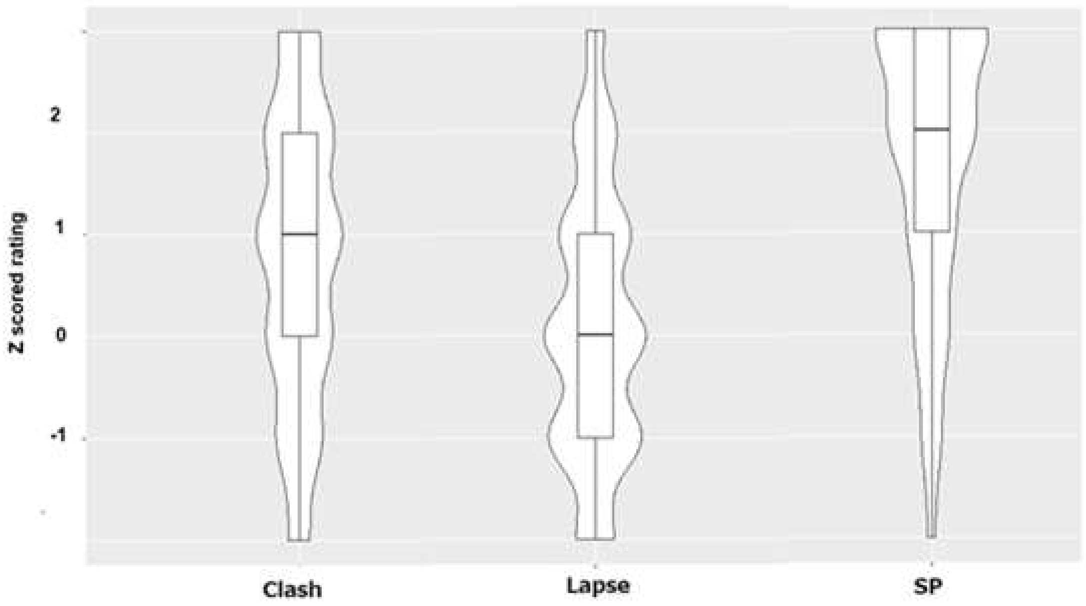


Figure 6.7: Violin plots of z-scores, shown separately for each language. Data from linguistic stimuli only. Data from musicians was excluded.

From Figure 6.7, it is clear that the rating of SP (mean = 1.397, SE = 0.071) was highest, while the second highest was clash (mean = 0.7, SE = 0.074). The rating of lapse was lowest (mean = 0.116, SE = 0.065). In other words, when violations were due to a lapse, listeners rated the differences from the familiarisation stimuli as less strong than when the violations were due to a clash. Structure-preserving stimuli showed higher ratings.

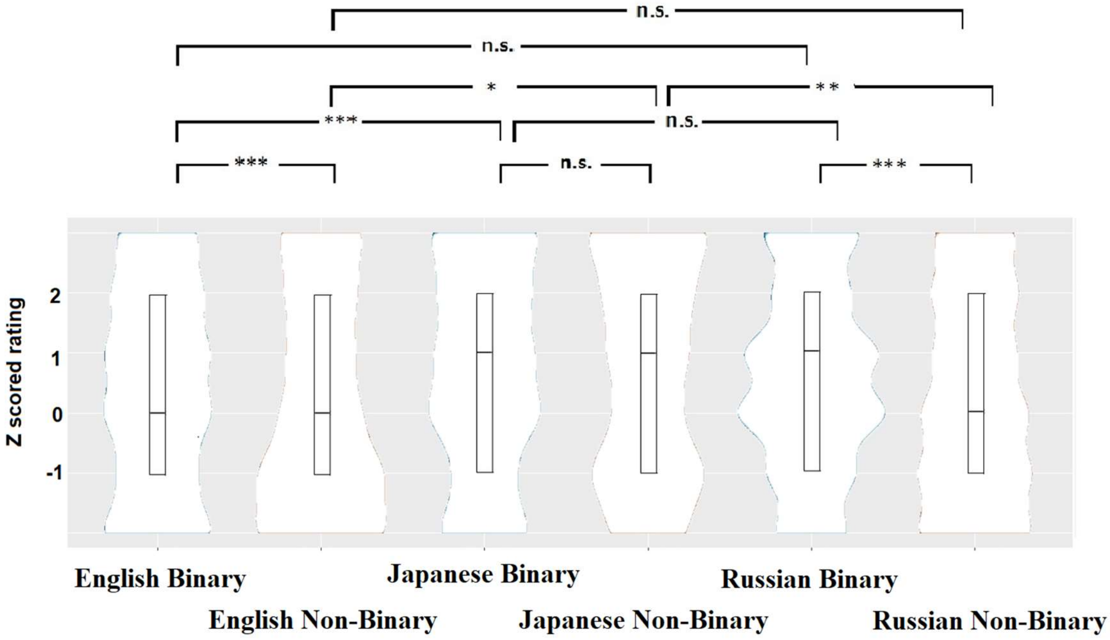


Figure 6.8: Violin plots of z-scores, shown separately for interaction between language and rhythm type. Data from linguistic stimuli only. Data from musicians was excluded.

Figure 6.8 and Table 6.4 show that there were no differences between English and Russian speakers in either binary rhythm (English mean = 0.257, SE = 0.031; Russian mean = 0.403, SE = 0.033) or non-binary rhythm (English mean = 0.112, SE = 0.031; Russian mean = 0.134, SE = 0.03). However, ratings of binary rhythm (mean = 0.33, SE = 0.035) and non-binary rhythm (mean = 0.332, SE = 0.034) from Japanese speakers were higher than those of English speakers. There was no statistical difference between Japanese and Russian ratings for binary rhythm, but the Japanese rating for non-binary rhythm was higher than that of Russian speakers.

Japanese ratings of binary rhythm were not statistically distinct from their ratings for non-binary rhythm. On the other hand, ratings of binary rhythm from both English and Russian speakers were higher than non-binary rhythm: in other words, for Japanese participants, violations in both rhythm types were considered comparable, while for English and Russian speakers, violations in binary rhythm led to differences that were perceived to be larger.

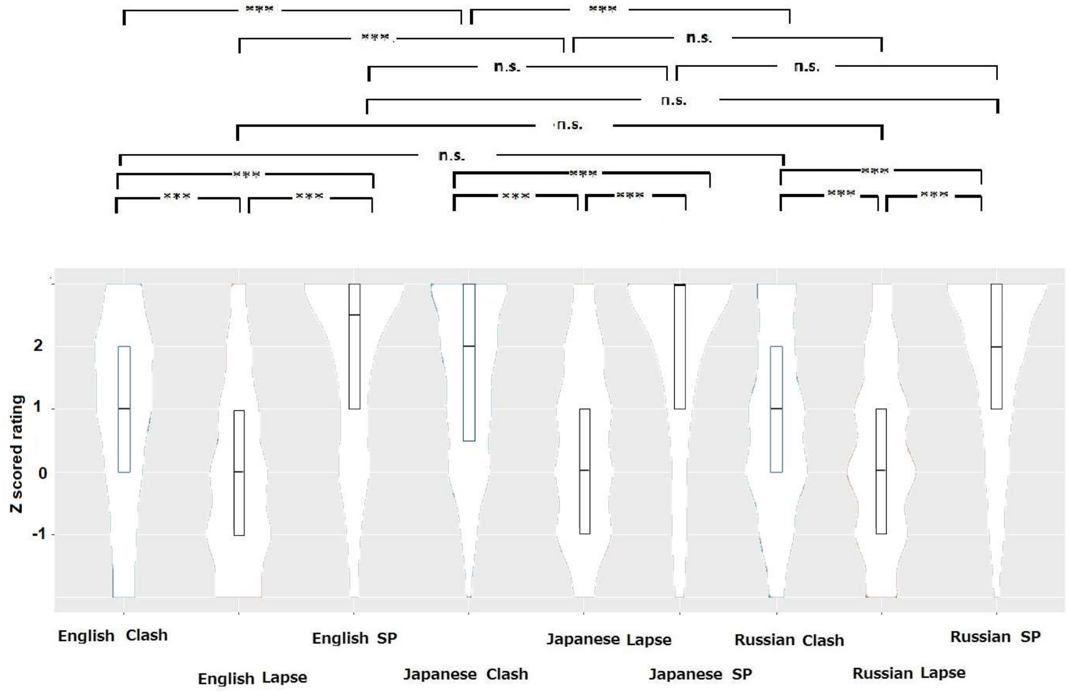


Figure 6.9: Violin plots of z-scores, shown separately for interaction between language and violation type. Data from linguistic stimuli only. Data from musicians were excluded.

Looking at Figure 6.9 and Table 6.4, it is clear that the rating for SP (structure-preserving stimulus) is the highest and the second highest is the clash, while the rating for lapse is relatively low, which means that the lapse was tolerated. The reason for the high rating of SP would be due to the fact that participants were asked to rate rhythmic difference, not metre difference.

Figure 6.9 shows that there was no statistical difference between English and Russian speakers in all violation types (clash, lapse, and SP): English clash mean rating = 0.585, SE = 0.045, English lapse mean rating = -0.029, SE = 0.0716, and English sp mean rating = 1.222, SE = 0.102, versus Russian clash mean rating = 0.551, SE = 0.066, Russian lapse mean rating = 0.155, SE = 0.053, Russian sp mean rating =

1.304, SE = 0.092. It also shows that Japanese ratings for the clash (mean rating = 0.869, SE = 0.064) were higher than those of Russian speakers, but the Japanese ratings for lapse (mean rating = 0.17, SE = 0.05) and SP (mean rating 1.219, SE = 0.111) were statistically identical to Russian ratings for lapse and SP, respectively. Looking at the difference between English and Japanese ratings, Japanese ratings of clash and lapse were higher than those of English clash and lapse, but there was no difference in SP.

6.3.3 Summary of the results of linguistic stimuli

With respect to the results of accuracy and ratings of linguistic stimuli, the only difference between English and Russian speakers was that Russian accuracy of non-binary rhythm was higher than that of English speakers. Japanese ratings and accuracy were mostly higher than English speakers, except for SP, for which ratings were not statistically different. The relationship between Japanese and Russian speakers was complicated: Japanese ratings of lapse, SP, and binary rhythm, and their accuracies of binary rhythm were identical, while Japanese ratings and accuracies in the other areas were higher than those of the Russian speakers. The results seem to demonstrate the influence of native language, because English and Russian, whose rhythms in native language have similar regularity, had similar tendencies in accuracy and rating.

6.4 Musical Stimuli

6.4.1 Accuracy

Table 6.5 presents the results obtained from LMM for the accuracy of musical stimuli. The main effects of language, rhythm type, and violation type were significant, and there was an interaction among language and rhythm type. However, there was no interaction between language and violation type (see Table 6.5).

Table 6.5 Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Language	**2, 92**	**6.47**	**0.0023**
Rhythm type	**1, 848**	**431.29**	**< 0.0001**
Violation type	**2, 848**	**237.88**	**< 0.0001**
Native_language*Rhythm type	**2, 848**	**11.46**	**< 0.0001**
Native_language*Violation type	4, 840	0.4	n.s.

Table 6.6: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
English binary vs. Japanese binary	**-0.5**	**120. 1**	**-2.85**	**0.005**
English binary vs. Russian binary	-0.3	120. 1	-1.43	n.s.
English binary vs. English non-binary	**0.9**	**848. 0**	**10.42**	**<0.0001**
Japanese binary vs. Russian binary	0.3	120. 1	1.39	n.s.
Japanese binary vs. Japanese non-binary	**0.9**	**848. 0**	**9.75**	**<0.0001**
Russian binary vs. Russian non-binary	**1.4**	**848. 0**	**15.76**	**<0.0001**
English non-binary vs. Japanese non-binary	**-0.5**	**120. 1**	**-3.03**	**0.003**
English non-binary vs. Russian non-binary	0.3	120. 1	1.44	n.s.
Japanese non-binary vs. Russian non-binary	**0.8**	**120. 1**	**4.41**	**<0.0001**
Clash vs. Lapse	**1.3**	**848. 0**	**15. 52**	**<0.0001**
Clash vs. SP	**0.712**	**848. 0**	**4.01**	**<0.0001**
Lapse vs. SP	**-0.725**	**848. 0**	**-4.21**	**<0.0001**

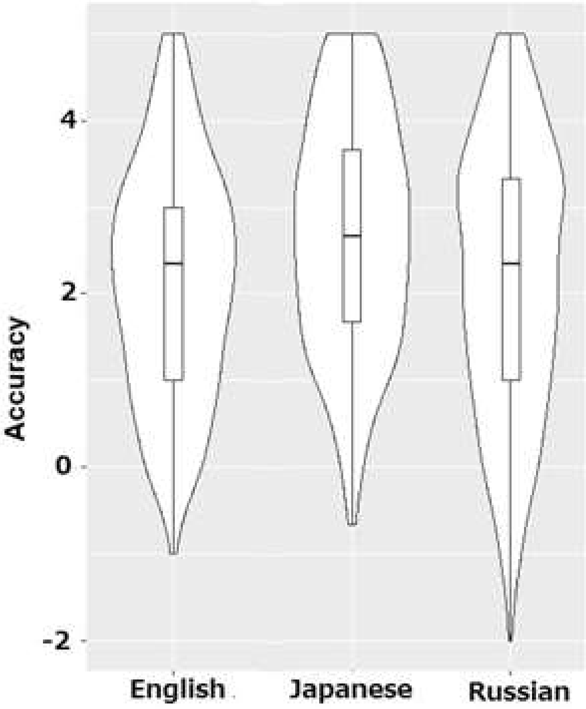


Figure 6.10: Violin plots of accuracy, shown separately for each language. Data from musical stimuli only. Data from musicians was excluded.

Figure 6.10 shows that the rating of Japanese (mean = 2.684, SE = 0.121) was higher than those of English (mean = 2.153, SE = 0.117) and Russian (mean = 2.152, SE = 0.121) speakers. This tendency of the English and Russian groups, whose native languages have similar rhythm, was identical to that shown in the previous results.

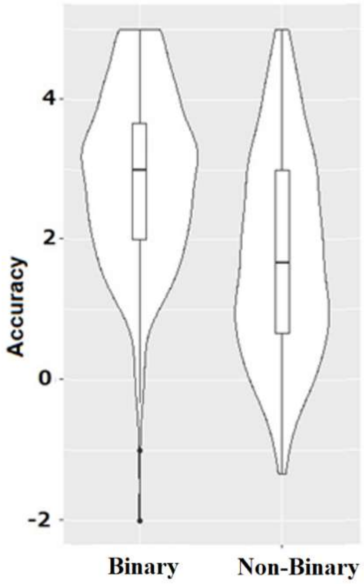


Figure 6.11: Violin plots of accuracy. Data from musical stimuli only. Data from musicians was excluded.

It is clear from Figure 6.11 that the accuracy of binary rhythm (mean = 2.876, SE = 0.074) was higher than non-binary rhythm (mean = 1.784, SE = 0.074). This seems to be a reasonable result, as the accuracy of comparison in the easy (binary rhythm trials) task was higher than in the more difficult (non-binary rhythm trials) task.

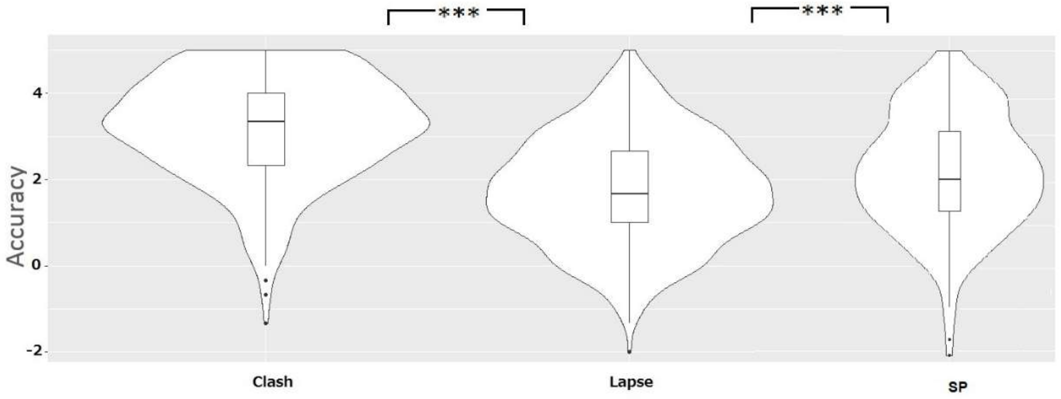


Figure 6.12: Violin plots of accuracy, shown separately for each violation type. Data from musical stimuli only. Data from musicians was excluded.

Figure 6.12 and Table 6.6 show that the accuracy with which participants detected clashes (mean accuracy = 3.191, SE = 0.062) was higher than their accuracy for lapses (mean accuracy = 1.754, SE = 0.052), similar to the results for linguistic stimuli (see Section 2.3.1). Accuracy of the SP fell between clash and lapse (mean = 2.479, SE = 0.087).

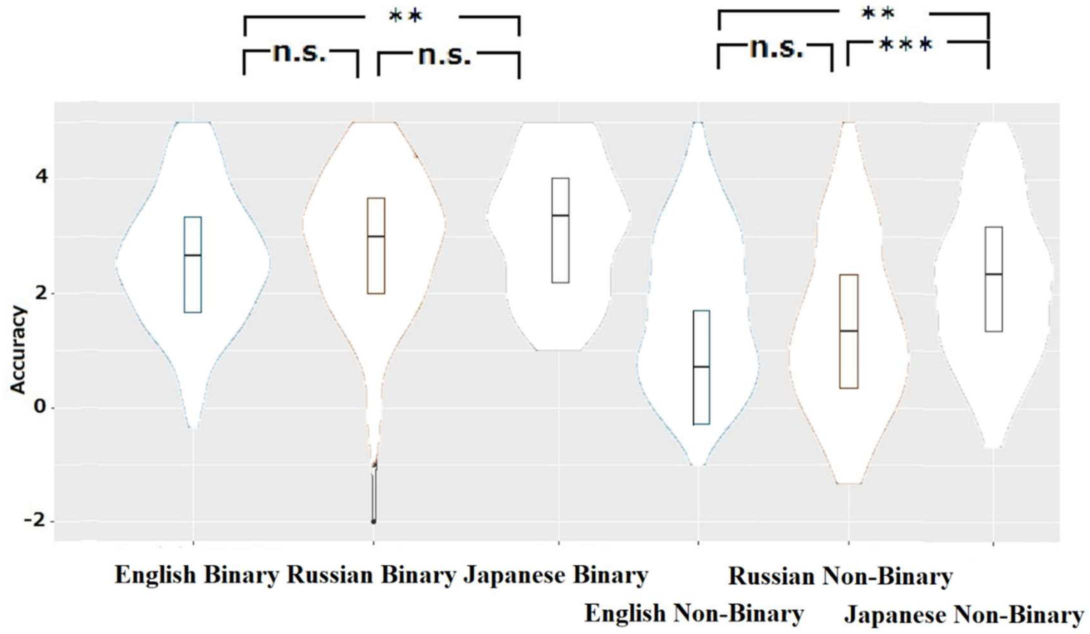


Figure 6.13: Violin plots of accuracy, shown separately for interaction between language and rhythm type. Data from musical stimuli only. Data from musicians was excluded.

Looking at Figure 6.13 and Table 6.6, it is clear that the accuracy of binary rhythm is higher than the accuracy of non-binary rhythm, regardless of participants' native language.

It is apparent from these figures that there was no difference between English and Russian speakers, both in binary and non-binary rhythm (English mean accuracy of binary rhythm = 2.618, SE = 0.09, English mean accuracy of non-binary rhythm = 1.688, SE = 0.104, versus Russian mean accuracy of binary rhythm = 2.883, SE = 0.103, Russian mean accuracy of non-binary rhythm = 1.447, SE = 0.116). Japanese accuracies (Japanese mean accuracy of binary rhythm = 3.133, SE = 0.091, Japanese mean accuracy of non-binary rhythm = 2.236, SE = 0.11) were higher than English both in a binary and non-binary rhythm. While Japanese accuracy in binary rhythm was not different from Russian speakers, Japanese accuracy of non-binary rhythm was higher than that of Russian speakers. These results were identical to the results of linguistic stimuli discussed in Section 6.3.

6.4.2 Rating

Table 6.7 is a summary of the best model output of the LMM for the z-scoresd rating of musical stimuli. As shown in Table 6.7, the main effects of language, rhythm type, and violation type were significant. There was also an interaction between language and the violation type, but there was no interaction between language and rhythm type.

Table 6.7: Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Language	**2, 83.34**	**4.79**	**0.0107**
Rhythm type	**1, 838.83**	**137.07**	**< 0.0001**
Violation type	**2, 840.18**	**129.53**	**< 0.0001**
Native_language*Rhythm type	2, 837.70	2.86	n.s.
Native_language*Violation type	**4, 839.93**	**6.36**	**< 0.0001**

Table 6.8: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
Binary vs. Non-binary	**0.5**	**838. 8**	**11.71**	**<0.0001**
English clash vs. Japanese clash	**-0.8**	**371. 0**	**-5.3**	**<0.0001**
English clash vs. Russian clash	-0.1	333. 6	-0.76	n.s.
English clash vs. English lapse	**1**	**836. 7**	**7.56**	**<0.0001**
English clash vs. English sp	**0.7**	**847. 6**	**5.05**	**<0.0001**
Japanese clash vs. Japanese lapse	**0.9**	**844. 2**	**6.06**	**<0.0001**
Japanese clash vs. Japanese sp	**0.7**	**845. 3**	**4.97**	**<0.0001**
Russian clash vs. Russian lapse	**1**	**832. 4**	**7.91**	**<0.0001**
Russian clash vs. Russian sp	**-0.7**	**832. 7**	**-5.98**	**<0.0001**
English lapse vs. Japanese lapse	**-0.8**	**389. 6**	**-5.1**	**<0.0001**
English lapse vs. Russian lapse	-0.2	334. 4	-1.07	n.s.
Japanese lapse vs. Russian lapse	0.2	347. 1	2.06	n.s.
English sp vs. Japanese sp	-0.1	334. 7	-0.55	n.s.
English sp vs. Russian sp	-0.2	303. 4	-1.67	n.s.
Japanese sp vs. Russian sp	**-0.8**	**326. 9**	**-4.82**	**<0.0001**

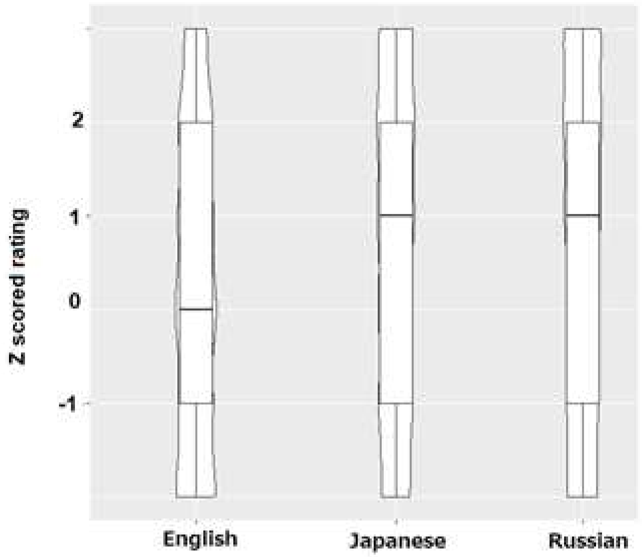


Figure 6.14: Violin plots of z-scores, shown separately for each language. Data from musical stimuli only. Data from musicians was excluded.

Figure 6.14 shows that the ratings of Japanese (mean = 0.718, SE = 0.069) and Russian (mean = 0.636, SE = 0.051) were statistically identical to each other, although the rating by English speakers (mean = 0.381, SE = 0.064) was lower than Japanese and Russian speakers. This result is different from the results in linguistic stimuli (accuracy and rating) and accuracy of musical stimuli, in the sense that English and Russian speakers were statistically different in Figure 6.14. A possible reason for this is that the difference in musical background between Japanese and Russians was small, while the English musical background was different to the other two groups.

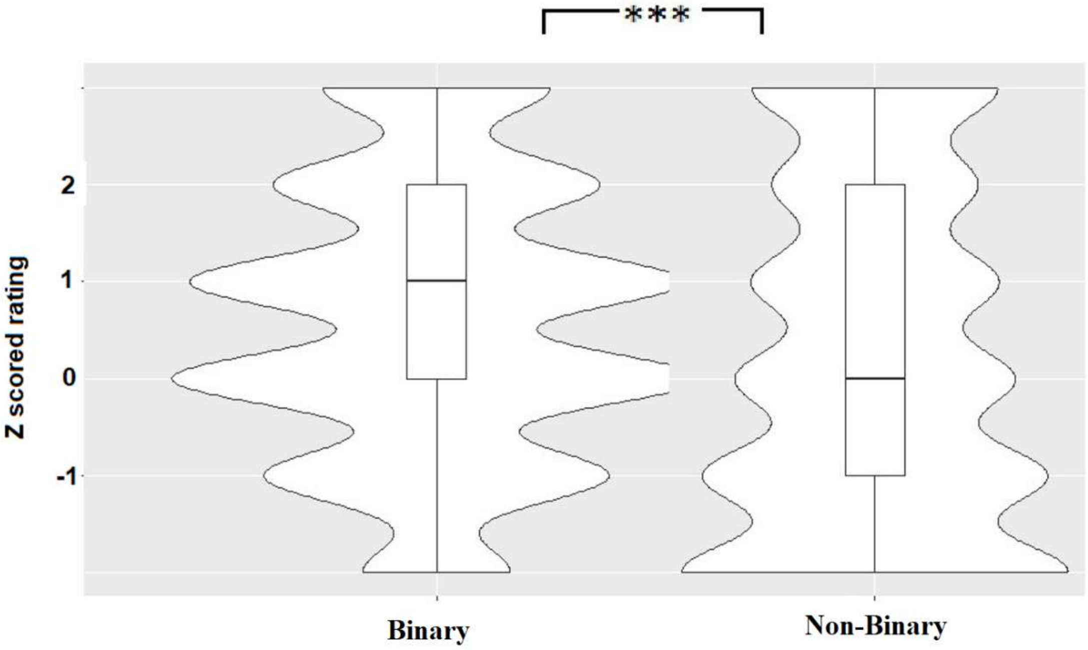


Figure 6.15: Violin plots of z-scores, shown separately for each language. Data from musical stimuli only. Data from musicians was excluded.

Figure 6.15 and Table 6.8 show that ratings for binary rhythm (mean rating = 0.517, SE = 0.04) were higher than ratings for non-binary rhythm (mean rating = 0.14, SE = 0.0392); in other words, rhythm violations were rated as stronger (i.e. worse) when they applied to stimuli with binary rhythm. There was no interaction between language and rhythm type.

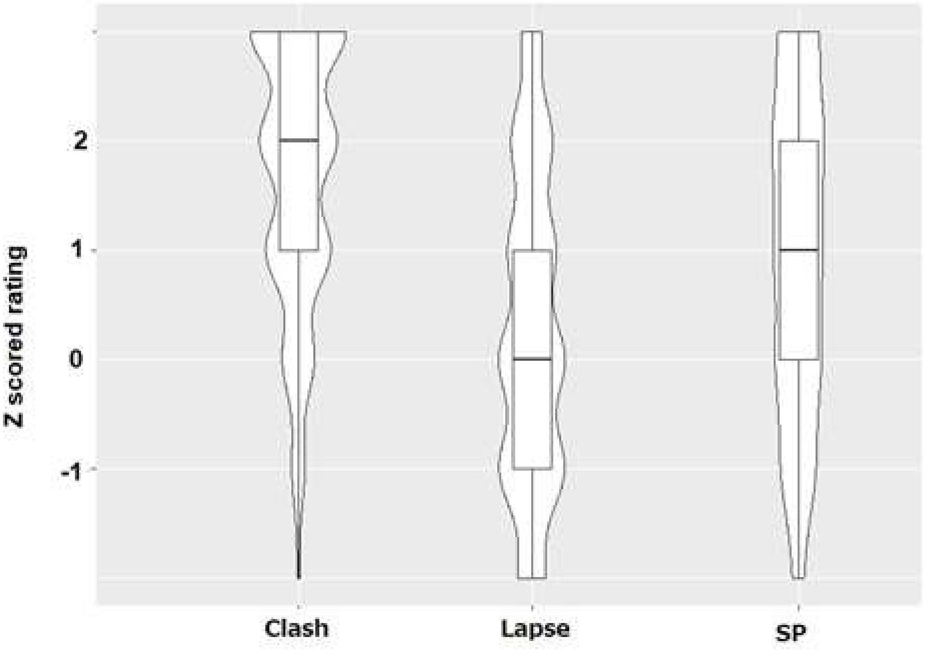


Figure 6.16: Violin plots of z-scores, shown separately for each violation type. Data from musical stimuli only. Data from musicians was excluded.

Figure 6.16 suggests that the rating of clash (mean = 1.253, SE = 0.055) was highest (i.e. clash was not tolerated) while the rating of lapse (mean = 0.135, SE = 0.055) was lowest. The rating of SP (mean = 0.612, SE = 0.054) was second highest; however, the result in linguistic stimuli suggested that the rating of SP (mean = 0.612, SE = 0.054) was highest.

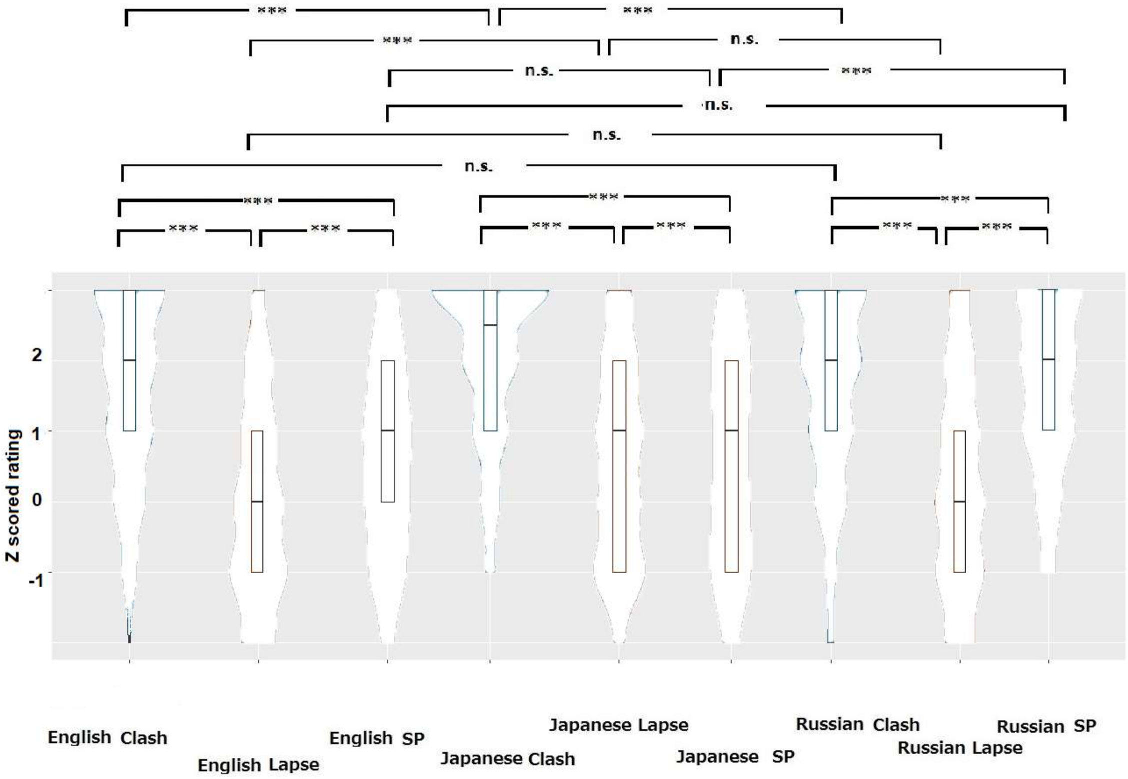


Figure 6.17: Violin plots of z-scores, shown separately for interaction between language and violation type. Data from musical stimuli only. Data from musicians was excluded.

Unlike the result for linguistic stimuli, the rating for SP was not the highest. Instead, the rating for clash was highest, while lapse (similar to the result for linguistic stimuli) was lowest.

It is apparent from Figure 6.17 and Table 6.8 that there was no difference between the ratings of English (English clash mean rating =1.038, SE = 0.061, English lapse mean rating = -0.021, SE = 0.055, English structure-preserving (SP) mean rating = 0.452, SE = 0.086) and Russian speakers (Russian clash mean rating = 1.197, SE = 0.061, Russian lapse mean rating = 0.125, SE = 0.062, Russian SP mean rating = 1.146, SE = 0.08). However, ratings by Japanese speakers for clash (Japanese clash mean rating = 1.335, SE = 0.059) were higher than English and Russian speakers. Also, ratings by Japanese speakers for lapse (Japanese lapse mean rating = 0.305, SE = 0.055) were higher than English speakers, but there was no difference between Japanese and Russian speakers. There was no difference in ratings for SP between English, Japanese (Japanese sp mean rating = 0.377, SE = 0.111), and Russian speakers.

6.4.3 Summary of the results of musical stimuli

Similar to the results of linguistic stimuli, there was no difference between English and Russian accuracies, and Japanese accuracies were always higher than those of English speakers. Moreover, like the results of linguistic stimuli, the Japanese accuracy of binary rhythm was not statistically different from Russian speakers, while Japanese accuracy of non-binary rhythm was higher than that of Russian speakers. Looking at the ratings, the Russian rating of SP was higher than that of Japanese speakers. Looking at the relationship between English and Japanese speakers, Japanese ratings were higher than English speakers, except for SP where there was no difference. To summarise, as in the results of linguistic stimuli, English and Russian speakers had a similar tendency, possibly indicating the influence of linguistic experience, as their linguistic rhythms are similar. Considering English and Japanese speakers, ratings and accuracies of Japanese speakers were higher than those of English speakers, except for SP where there was no difference between the two. Japanese ratings and accuracies tended to be higher than Russian speakers, but there was no statistical difference between them in binary rhythm or in accuracy and rating of lapse. Despite the high ratings and accuracies of Japanese speakers, the Japanese rating of SP was lower than that of Russian speakers. As with other results, the ratings and accuracy of clash were higher than lapse.

6.5 Pure tone Stimuli

6.5.1 Accuracy

There was no difference between the accuracy of long pure tone stimuli and short pure tone stimuli (linear mixed model, t = 0.875, df = 1.823e+03; p = n.s.); thus, the two types of pure tone stimuli were treated as a pure tone stimuli group and pooled for further analysis.

Table 6.9 below is a summary of the best model output of the LMM for the accuracy of pure tone stimuli. As shown in Table 6.9, the main effects of language, rhythm type, and violation type were significant. There was also an interaction between language and rhythm type, but there was no interaction between language and violation type.

Table 6.9: Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Language	**2, 92**	**4.77**	**0.0106**
Rhythm type	**1, 1798**	**498.47**	**< 0.0001**
Violation type	**2, 1798**	**274.7**	**< 0.0001**
Native_language*Rhythm type	**2, 1798**	**3.84**	**0.0215**
Native_language*Violation type	4, 1790	1.84	n.s.

Table 6.10: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
Clash vs. Lapse	**1.1**	**1798. 0**	**16.99**	**<0.0001**
Clash vs. SP	**0.792**	**1798. 0**	**3.59**	**<0.0001**
Lapse vs. SP	**-0.558**	**1798. 0**	**3.12**	**<0.0001**
English binary vs. Japanese binary	**-0.4**	**112. 7**	**-2.41**	**0.018**
English binary vs. Russian binary	-0.2	112. 7	-1.07	n.s.
Japanese binary vs. Russian binary	0.2	112. 7	1.32	n.s.
English non-binary vs. Japanese non-binary	**-0.5**	**112. 7**	**-3.19**	**0.002**
English non-binary vs. Russian non-binary	0	112. 7	-0.1	n.s.
Japanese non-binary vs. Russian non-binary	**0.5**	**112. 7**	**3.05**	**0.003**

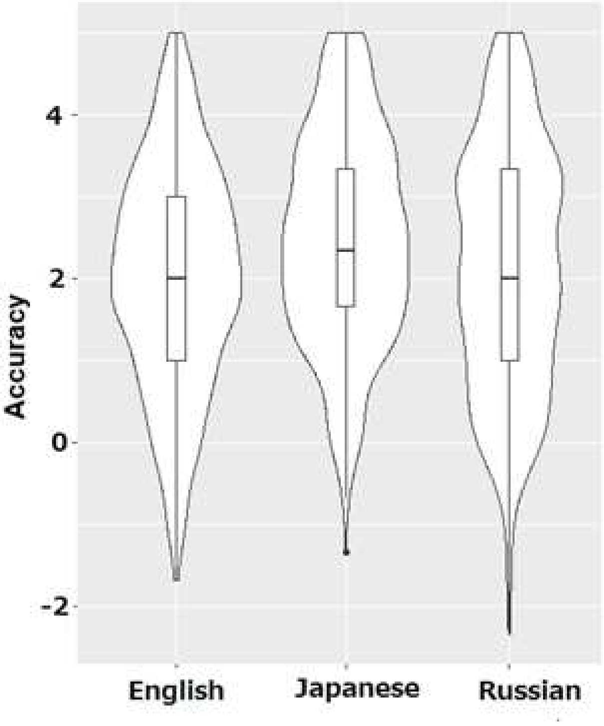


Figure 6.18: Violin plots of accuracy, shown separately for each language. Data from tonal stimuli only. Data from musicians was excluded.

Figure 6.18 shows that the accuracy of Japanese speakers (mean = 2.478, SE = 0.113) was higher than that of English (mean = 2.014, SE = 0.109) and Russian (mean = 2.11, SE = 0.113), while there was no statistical difference between English and Russian speakers. The result is identical to the two previous results of accuracy in linguistic and musical stimuli.

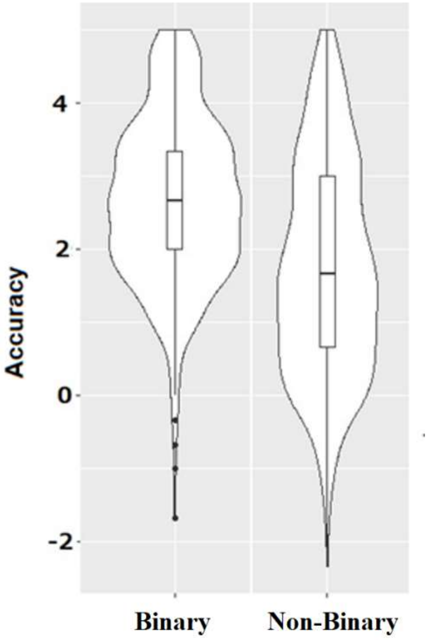


Figure 6.19: Violin plots of accuracy, shown separately for each rhythm type. Data from tonal stimuli only. Data from musicians was excluded.

From Figure 6.19, it is clear that the accuracy of binary rhythm (mean = 2.674, SE = 0.068) was higher than non-binary rhythm (mean = 1.728, SE = 0.068). This result is identical to the results in linguistic and musical stimuli. This clear tendency is convincing: the accuracy of binary rhythm is higher than the non-binary rhythm.

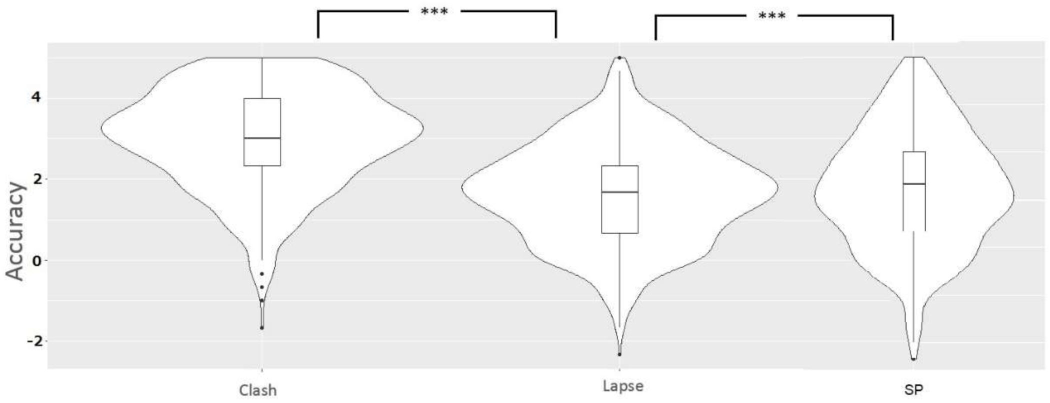


Figure 6.20: Violin plots of accuracy, shown separately for each violation type. Data from tonal stimuli only. Data from musicians was excluded.

From Figure 6.20 and Table 6.10, it is clear that the accuracy of clash (mean accuracy = 3.008, SE = 0.044) was higher than lapse (mean accuracy = 1.658, SE = 0.036). The result is identical to that of linguistic and musical stimuli. Accuracy of the SP fell between clash and lapse (mean = 2.216, SE = 0.086).

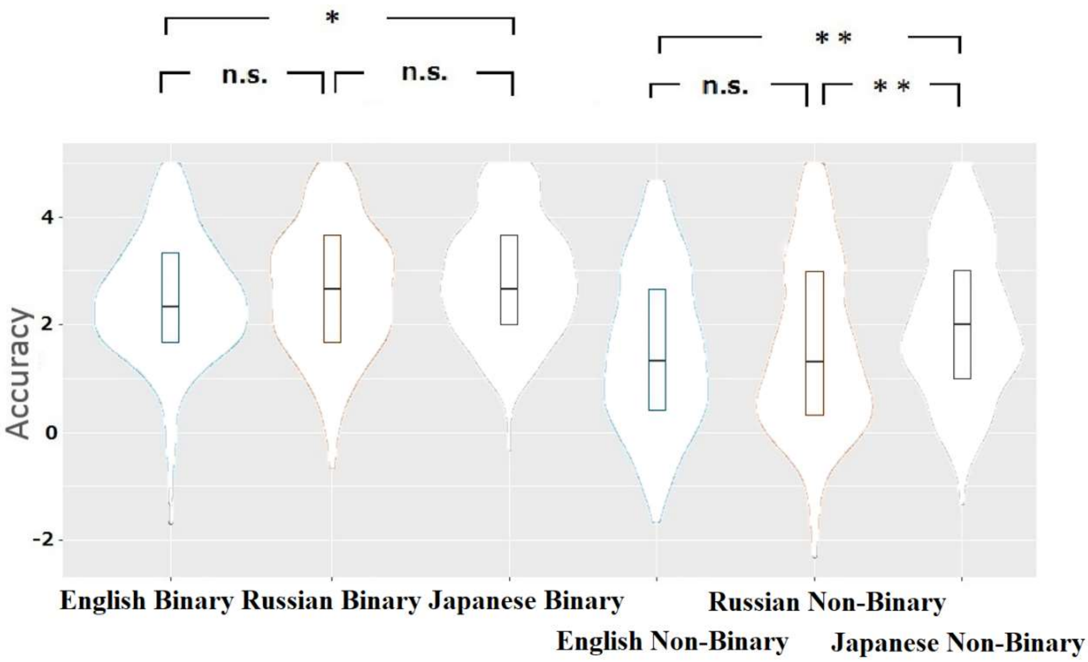


Figure 6.21: Violin plots of accuracy, shown separately for interaction between language and rhythm type. Data from tonal stimuli only. Data from musicians was excluded.

Figure 6.21 and Table 6.10 show that there were no differences between English (English mean accuracy of binary rhythm = 2.481, SE = 0.041, English mean accuracy of non-binary rhythm = 1.546, SE = 0.078) and Russian (Russian mean accuracy of binary rhythm = 2.676, SE = 0.065, Russian mean accuracy of non-binary rhythm = 1.559, SE = 0.086) speakers in binary and non-binary rhythm. While Japanese accuracy of binary rhythm (Japanese mean accuracy of binary rhythm = 2.881, SE = 0.062) was not different from that of Russian speakers, Japanese accuracy of non-binary rhythm (Japanese mean accuracy of non-binary rhythm = 2.075, SE = 0.077) was higher than that of Russian speakers. Japanese accuracies were higher than English speakers both in binary and non-binary rhythms.

These results are all identical to the results of musical stimuli; however, one result differs to those of the linguistic stimuli: English accuracy of non-binary rhythm was lower than Russian speakers.

6.5.2 Rating

There was no difference between the accuracy of long pure tone stimuli and short pure tone stimuli (linear mixed model, t = 1.063, df = 1.409e+04; p = n.s.); thus, the two types of pure tone stimuli were treated as a pure tone stimuli group and further analysed.

As shown in Table 6.11, the main effects of language, rhythm type, and violation type were significant. However, there was no interaction between language and rhythm type, or between language and violation type.

Table 6.11: Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Language	**2, 81.97**	**3.1948**	**0.0461**
Rhythm type	**1, 1798.90**	**183.3908**	**< 0.0001**
Violation type	**2, 1793.95**	**174.7183**	**< 0.0001**
Native_language*Rhythm type	2, 1797.97	2.4034	n.s.
Native_language*Violation type	4, 1782.09	1.6376	n.s.

Table 6.12: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
English vs. Japanese	**-0.2**	**84. 6**	**-2.43**	**0.017**
English vs. Russian	-0.1	79. 8	-0.56	n.s.
Japanese vs. Russian	0.2	81. 8	1.88	n.s.
Binary vs. Non-binary	**0.5**	**1798. 9**	**13.54**	**<0.0001**
Clash vs. Lapse	**1**	**1792. 7**	**16.05**	**<0.0001**
Clash vs. SP	**0.4**	**1792. 6**	**6.41**	**<0.0001**
Lapse vs. SP	**-0.6**	**1797. 1**	**-9.84**	**<0.0001**

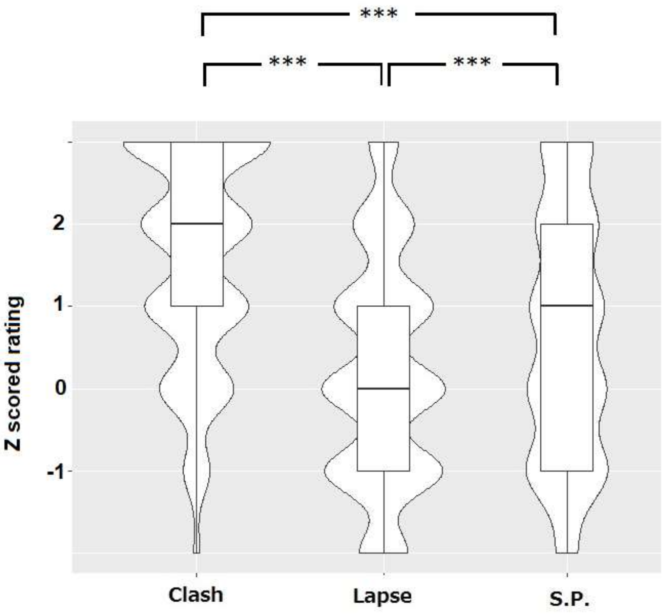


Figure 6.22: Violin plots of z-scores, shown separately for each violation type. Data from tonal stimuli only. Data from musicians was excluded.

Figure 6.22 and Table 6.12 show that the rating of clash (clash mean accuracy = 1.115, SE = 0.025) was highest and the second highest rating was for lapse (lapse mean accuracy = 0.123, SE = 0.022), while the rating of SP (sp mean accuracy = 0.504, SE = 0.043) was lowest. Except for the result of linguistic stimuli, in which SP was highest, the rating of SP had a tendency to be the second highest.

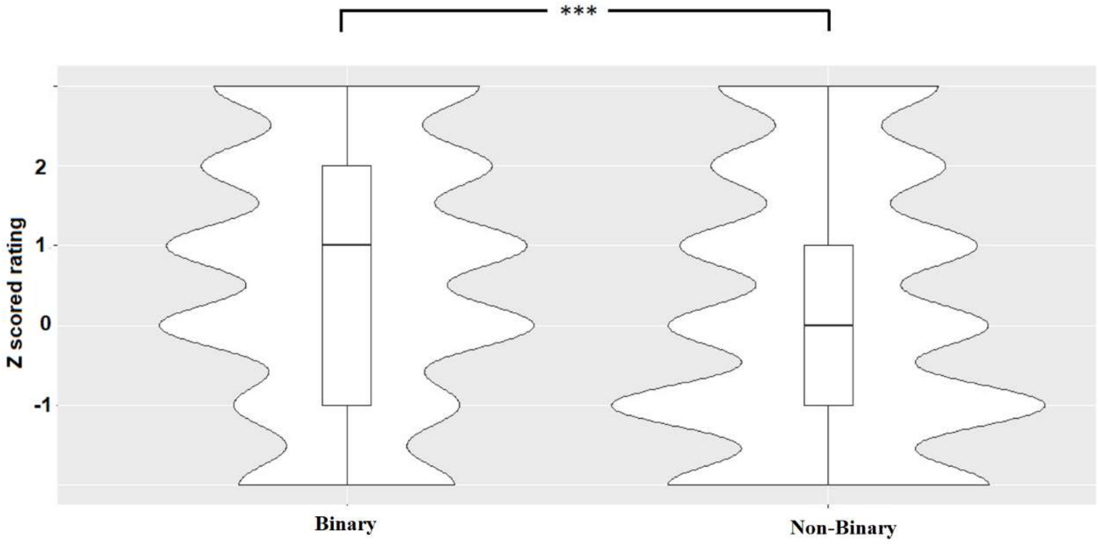


Figure 6.23: Violin plots of z-scores, shown separately for each rhythm type. Data from tonal stimuli only. Data from musicians was excluded.

Figure 6.23 and Table 6.12 show that the rating of binary rhythm (mean = 0.46, SE = 0.035) was higher than that of non-binary rhythm (mean = 0.117, SE = 0.027). Except for an exceptional case, in which ratings by musicians and Japanese speakers for binary rhythm did not differ from non-binary rhythm, the rating for binary rhythm was higher than non-binary rhythm throughout the results.

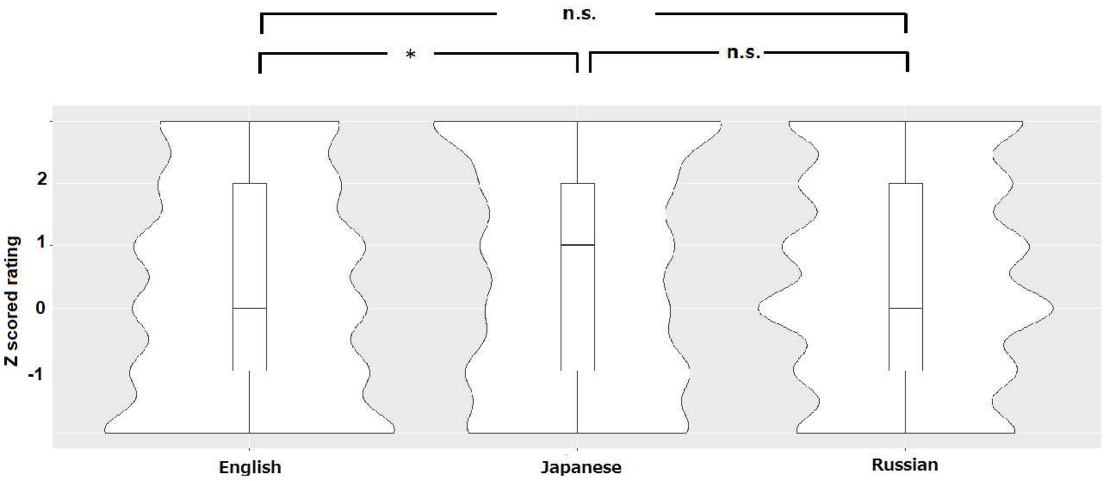


Figure 6.24: Violin plots of z-scores, shown separately for each language. Data from tonal stimuli only. Data from musicians was excluded.

Table 6.24 and Table 6.12 suggest that there was no difference between the ratings of Japanese (Japanese mean rating = 0.319, SE = 0.034) and Russian speakers (Russian mean rating = 0.323, SE = 0.033) or between the ratings of English (English mean rating = 0.226, SE = 0.032) and Russian speakers. The only statistical difference was found between Japanese and English speakers. Looking at the median and quartile, there is no difference between English and Russian speakers while Japanese were higher than both English and Russian speakers. This tendency that perceptions of English and Russian speakers are similar, while Japanese speakers are more sensitive to rhythm, was also found in other results.

6.5.3 Summary of the results of tonal stimuli

Considering the accuracy, as shown in the previous two sections, Japanese accuracies were higher than those of English speakers, and there was no difference in accuracies between English and Russian speakers. While Japanese accuracy of non-binary rhythm was higher than that of Russian speakers, there was no difference between them in binary rhythm.

Looking at the ratings, there was no difference between English and Russian speakers, or between Japanese and Russian speakers. Japanese ratings were, regardless of the violation type, higher than those of English speakers. There was no difference in ratings between Japanese and Russian speakers, which differs from the results of musical and linguistic stimuli.

6.6 Summary of the results and discussion

As hypothesised at the beginning of this chapter in Hypothesis 1, all participants rated clash as a more significant violation than lapse. Japanese participants also accurately detected rhythmic violation of non-binary rhythm better than English and Russian speakers, as hypothesised in Hypothesis 2. Contrary to the idea that Japanese speakers will be more sensitive to lapse than English and Russian speakers, as suggested in Hypothesis 3, Japanese rating and accuracy of lapse was not significantly higher than those of English and Russian speakers. A conclusive consensus for Hypothesis 4 – that linguistic experience will affect the result of linguistic stimuli – was denied, as the influence of linguistic experience was not significantly clearer in the results of linguistic stimuli than in the results of musical and tonal stimuli. Details of the results will be discussed in the following sections.

6.6.1 Results of accuracy and discussion

With respect to the accuracy of the binary rhythm, Japanese accuracies were higher than those of English speakers in all stimulus types (linguistic, musical, and pure-tone stimuli). The accuracies of Russian speakers in binary rhythm did not differ from English and Japanese speakers in all stimulus types. In any case, Japanese accuracies of binary and non-binary rhythm were not lower than those of English and Russian speakers across all of the experiments.

Regarding accuracy of the non-binary rhythm, Japanese speakers were higher than English and Russian speakers in all stimulus types. Although the accuracy of English and Russian speakers for non-binary rhythm did not differ statistically in musical and pure-tone stimuli, English speakers were lower Russian speakers in linguistic stimuli.

The results of accuracy for binary and non-binary rhythm can be summarised as follows: accuracies of Japanese speakers were higher than those of English speakers, while accuracies of Russian speakers fell between Japanese and English speakers, and accuracies of English and Russian speakers were similar. As mentioned in Chapter 1, Russian musical rhythm and their musical education are like those of Japan, while Russian language rhythm is similar to English language rhythm. Therefore, while it seems that linguistic experience affects rhythm perception, the result may suggest that there are correlations between linguistic and musical experiences on perception. Considering the teaching hours of music in England, Japan, and Russia (see Section 2.1.1), another possibility is that it is only musical experience that affects the result, as the accuracies of each country correspond with their total teaching hours. These ideas will be discussed further in the next chapter, which is a comparison of the results of musicians and non-musicians, because it is impossible to conclude if both the linguistic and musical experiences affect the rhythmic accuracies of binary and non-binary rhythm without first comparing musician and non-musician groups.

For the accuracy of each violation type (lapse, clash, and SP), there were no interactions among native language and violation type across all stimulus types (linguistic, musical, and pure tone stimuli). This suggests that the linguistic experience did not affect the accuracy of each violation type, although it was hypothesised that Japanese speakers would perceive the lapse differently due to the frequent use of lapse in their native language (see Section 2.2.7).

Looking at the main effect, the important finding is that the accuracy of Japanese speakers was higher than English and Russian speakers, regardless of the type of stimuli (linguistic, musical, or tonal), while there was no statistical difference between English and Russian speakers, whose linguistic rhythms have the identical rhythmic characteristics (relatively constant alternation of stressed and unstressed syllables). These results suggest that linguistic experience affects the accuracy of rhythmic accuracy, while English and Russian speakers have different musical experiences and backgrounds.

The difference between English and Russian speakers was limited to the accuracy of non-binary rhythm in linguistic stimuli. A possible reason for this is a difference in cues between English and Russian. The least important of the cues to detect accent in English is pitch (Kochanski, Grabe, Coleman, & Rosner, 2005), and pitch was used to indicate accent in the linguistic stimuli.

6.6.2 Results of rating and discussion

Concerning the ratings (z-scores) of each rhythm type (binary and non-binary), ratings by English speakers were not statistically different from Russian speakers across all stimulus types (linguistic, musical, and pure tone stimuli). Japanese ratings were higher (they rated stimuli with rhythmic violations as more dissimilar to controls) than English and Russian speakers across all stimulus types, except for ratings for binary rhythm in linguistic stimuli, in which the Japanese rating was statistically identical to that of the Russian speakers. Looking at the rating for each violation type, as in all the previous results, the rating for English speakers was identical to that for Russian speakers across all stimulus types. Japanese ratings for each violation type were higher than those of English speakers, except for in linguistic and musical SP, in which there was no statistical difference between Japanese and English speakers. Similarly, Japanese ratings for each violation type tended to be higher than those of Russian speakers, but in pure tone stimuli, there was no difference between the Japanese and Russian groups.

Moreover, while Japanese ratings for clash were higher than Russian speakers in linguistic and musical stimuli, there was no statistical difference in the rating for lapse between Japanese and Russian speakers in linguistic and musical stimuli. For SP, the Russian rating of SP was even higher than the Japanese rating of SP, and there was no statistical difference between the Japanese rating of SP and that of Russian speakers in the linguistic stimuli.

As Gouskova and Roon (2013) suggest, Russians tolerate lapse in their language similarly to Japanese speakers, and this might explain why there was a similarity between Japanese and Russian groups. Interestingly, when considering the Japanese rating for SP, statistical results, distribution shape, and median show that the Japanese rating for SP was lower than English and Russian speakers, while other Japanese ratings and accuracies were higher than the English and Russian groups. One reason for the low Japanese rating for the SP could be that Japanese speakers count rhythm based on mora (i.e. the Japanese long vowel is counted as two units), this would mean that a lengthened single element in SP was counted as two elements. Therefore, the difference between the familiarisation stimulus and SP, where two elements in SP are unified into a long sound, creating an element twice as long as the other elements, could be perceived as identical in both stimuli by Japanese speakers (see Section 5.3 on the difference between familiarisation and SP stimuli).

Similar to the tendency of Japanese speakers to perceive the succession of linguistic units as a succession of a constant moraic unit (Hattori, 1980), Japanese speakers count non-syllabic morae that are written into individual letters as a single mora.

Another important finding concerns the main effect of linguistic experience, and where the rating of musical stimuli was different from other results of accuracies and ratings in different sound types (linguistic and tonal stimuli): there was no difference between Japanese and Russian speakers, while the English rating was lower than Japanese and Russian speakers. This possibly denies the influence of native language, as Japanese and Russian languages have different rhythms. A possible reason for this might be the influence of musical background, in which Japanese and Russian speakers are familiar with the non-binary rhythm in their traditional music. This influence of the musical experience will be further discussed in the next chapter.

In conclusion, in most cases, including both rating and accuracy, the rhythmic perception of English and Russian participants responded in a similar manner to the stimuli and were accurate to a comparable degree. This is likely to be a result of similarities in the speech rhythms of Russian and English, as both are based on an even alternation of prominent and non-prominent elements (stressed and unstressed syllables). On the other hand, musical backgrounds and the musical rhythms prevailing in Russia and the UK differ. Therefore, the similarities between the responses of the English and Russian participants suggest that differences in musical experience do not affect the rhythm perception.

Japanese participants were more accurate in detecting rhythm violations and rated such violations more highly (i.e. thought them worse than the English and Russian participants did). This sensitivity of the Japanese groups is possibly due to Japanese speakers being familiar with the non-binary rhythms used as test stimuli.

Chapter 7: The role of musical experience: English and Japanese musicians and non-musicians

7.1 Introduction

This chapter focuses on whether, and how, musical experience, both in terms of practice and familiarity with different types of rhythm, affects rhythm perception. Furthermore, this chapter addresses an additional research question, namely whether the musical experience is as important as linguistic experience. In order to address these questions, a comparison is made here between English musicians and non-musicians on the one hand, and Japanese musicians and non-musicians on the other. The following hypotheses, which were demonstrated in the earlier method chapter, are also tested in this chapter:

Hypothesis 1: Musicians will perceive musical rhythm differently from linguistic and tonal rhythms.

Hypothesis 2: Musical experience will lead to greater accuracy in perceiving rhythm violations and to higher difference ratings among musicians in relation to non-musicians. This should apply independently of the native language of participants.

Hypothesis 3: It is hypothesised that linguistic experience will affect rhythm perception more than musical experience, as we learn language first and spend more time on language than music.

7.2 Statistical method

Statistical methods are similar to those of linguistic experience, but one main difference is that there is no level of factor for Russian, while musical experience (musician or non-musician) is added as a fixed factor in this chapter (for more details, see Section 5.7).

Although there are two types of tonal stimuli (long and short), it was at first analysed whether the difference of tonal length affects the rating and accuracy individually. The detail of the analysis for the tonal stimuli can be found in the sections for tonal stimuli.

The accuracy of the linguistic stimuli was statistically examined through R (Dependent variable= accuracy, Random factor = participant. Fixed factors = native language [English, Japanese], musical experience [musician, non-musician], rhythm type [binary, non-binary], and violation type [clash, lapse, SP]). The script 7.1 below was used for these statistical analyses to examine accuracies of linguistic stimuli.

(7.1) Script: mb4 = lmer(accuracy ~ native_language*musical_experience + musical_experience*violation_type + musical_experience*rhythm_type + (1|participant), data = accuracy_linguisticstimuli_musicalexperiencecomparison)

The rating of the linguistic stimuli was statistically examined through R (Dependent variable= z-scoresd rating, Random factor = participant, Fixed factors = native language [English, Japanese], musical experience [musician, non-musician], rhythm type [binary or non-binary], musical experience [musician and non-musician], and violation type [SP, clash, lapse). The script 7.2 below was used for these statistical analyses to examine ratings of linguistic stimuli.

(7.2) Script:

```
mb5 = lmer(zaverage ~ native_language*musical_experience +
musical_experience*violation_type + musical_experience*rhythm_type (1|participant),
data=rating_linguisticstimuli_musicalexperiencecomparisongroup
```

The accuracy of the musical stimuli was statistically examined through R (Dependent variable= accuracy, Random factor = participant, Fixed factors = native language [English, Japanese], musical experience [musician or non-musician], rhythm type [binary or non-binary], musical experience [musician and non-musician], and violation type [clash, lapse, sp]). The script 7.3 below was used for these statistical analyses to examine accuracies of musical stimuli.

(7.3) Script:

```
ma4 = lmer(accuracy ~ native_language*musical_experience +
musical_experience*violation_type + musical_experience*rhythm_type + (1|participant),
data=accuracy_musicalstimuli_musicalexperiencecomparisongroup)
```

The rating of the musical stimuli was statistically examined through R (Dependent variable= z-scoresd rating, Random factor = participant, Fixed factors = native language [English, Japanese], musical experience [musician, non-musician], rhythm type [binary or non-binary], musical experience [musician and non-musician], and violation type [clash, lapse, SP]). The script 7.4 below was used for these statistical analyses to examine ratings of musical stimuli.

(7.4) Script:

```
ma5 = lmer(zaverage ~ native_language*musical_experience +
musical_experience*violation_type + musical_experience*rhythm_type + (1|participant),
data=rating_musicalstimuli_musicalexperiencecomparisongroup)
```

The accuracy of the tonal stimuli was statistically examined through R (Dependent variable= accuracy, Random factor = participant, Fixed factors = native language [English, Japanese], musical experience [musician, non-musician], rhythm type [binary or non-binary], musical experience [musician and non-musician], and violation type [clash, lapse, SP]). The script 7.5 below was used for these statistical analyses to examine accuracies of tonal stimuli.

(7.5) Script:

```
ma6 = lmer(accuracy ~ native_language*musical_experience +
musical_experience*violation_type + musical_experience*rhythm_type
+ (1|participant), data=accuracy_musicalstimuli_musicalexperiencecomparisongroup)
```

The rating of the tonal stimuli was statistically examined through R (Dependent variable = z-scoresd rating, Random factor = participant, Fixed factors = native language [English, Japanese], musical experience [musician, non-musician], rhythm type [binary or non-binary], musical experience [musician and non-musician], and violation type [clash, lapse, SP]). The script 7.6 below was used for these statistical analyses to examine ratings of tonal stimuli.

(7.6) Script:

```
m5c = lmer(zaverage ~ native_language*musical_experience +
musical_experience*violation_type + musical_experience*rhythm_type + (1|participant),
data=rating_puretonestimuli_musicalexperiencecomparisongroup)
```

7.3 Result of linguistic stimuli

7.3.1 Accuracy

As shown in Table 7.1, the best model output of the LMM for the accuracy of linguistic stimuli showed effects of native language and interactions between musical experience and violation type and between musical experience and rhythm type. However, the best model output did not reveal an effect of musical experience.

Table 7.1: Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Musical experience*Language	1, 118.61	1.717	n.s.
Language	**1, 118.61**	**14.936**	**0.00018**
Musical experience*Violation type	**2, 1299.71**	**7.891**	**0.00039**
Musical experience*Rhythm type	**1, 1300.12**	**7.944**	**0.0048**

Table 7.2: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
Musician clash vs. Non-musician clash	0.298	254	2.376	n.s.
Musician clash vs. Musician lapse	**0.965**	**1300**	**11.726**	**< 0.0001**
Musician clash vs. Musician sp	-0.139	1300	-1.27	n.s.
Non-musician clash vs. Non-musician lapse	**0.832**	**1301**	**10.570**	**< 0.0001**
Non-musician clash vs. Non-musician sp	**-0.734**	**1299**	**-7.012**	**< 0.0001**
Musician lapse - Non-musician lapse	0.165	184	1.432	n.s.
Musician lapse vs. Musician sp	**-1.105**	**1300**	**-10.631**	**< 0.0001**
Non-musician lapse vs. Non-musician sp	**-1.567**	**1300**	**-15.824**	**< 0.0001**
Musician sp vs. Non-musician sp	-0.297	505	-1.945	n.s.
Musician binary vs. non-musician binary	-0.088	194	-0.76	n.s.
Musician binary vs. Musician non-binary	**0.46**	**1300**	**6.241**	**< 0.0001**
Non-musician binary vs. Non-musician non-binary	**0.748**	**1301**	**10.614**	**< 0.0001**
Musician non-binary vs. non-musician non-binary	0.199	199	1.692	n.s.

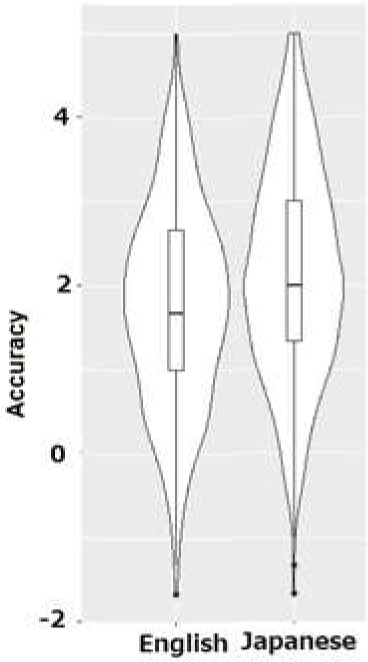


Figure 7.1: Violin plots (showing distribution shape, median, and interquartile range) of accuracy, shown separately for each language. Data from linguistic stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

From Figure 7.1, it is clear that the accuracy of Japanese speakers (mean = 2.14, SE = 0.08) was higher than English speakers (mean = 1.77, SE = 0.085) in linguistic stimuli, similar to the results in the previous chapter.

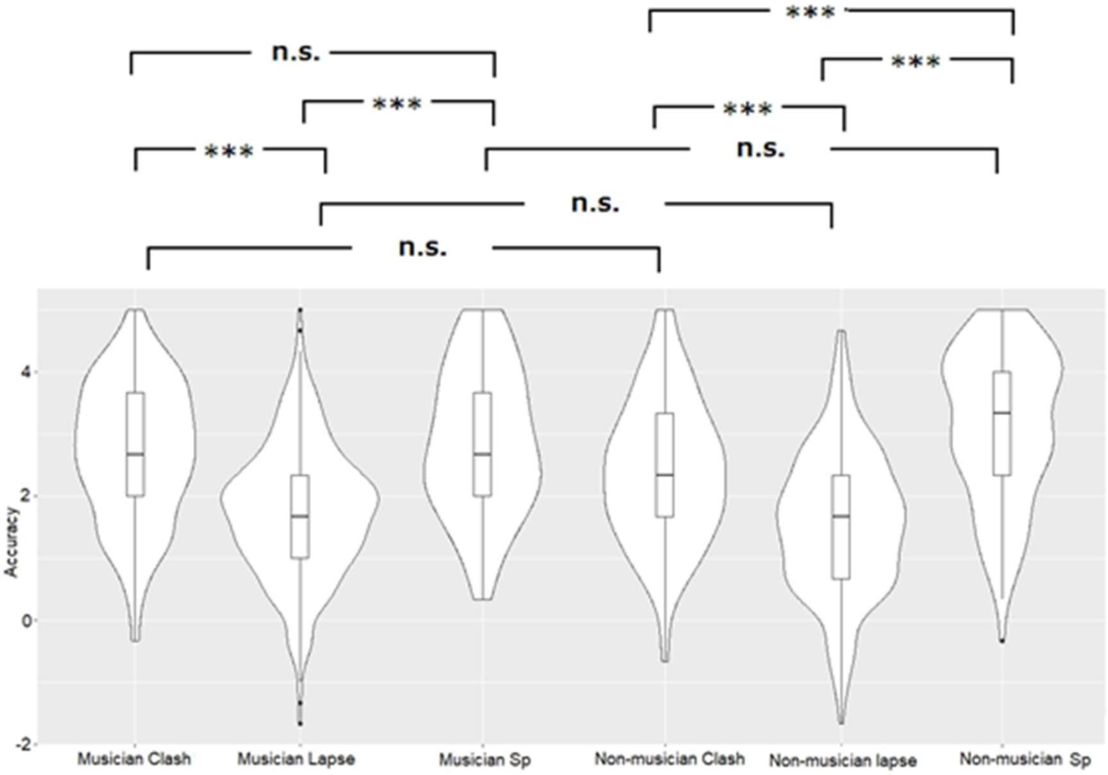


Figure 7.2: Violin plots of accuracy, shown separately for interaction between violation type and musical experience. Data from linguistic stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.2 and Table 7.2 show that there was no statistical difference between musicians and non-musicians in accuracies on each violation type (musicians: clash = 2.704 [SE = 0.078] lapse = 1.747 [SE = 0.081] SP = 2.846 [SE = 0.08], non-musicians: clash = 2.390 [SE = 0.079] lapse = 1.56 [SE = 0.076] SP = 3.115 [SE = 0.081]). The only difference between musician and non-musician was that there was no statistical difference between accuracy on clash and SP by musicians, while accuracy on SP by non-musicians was higher than accuracy for clash.

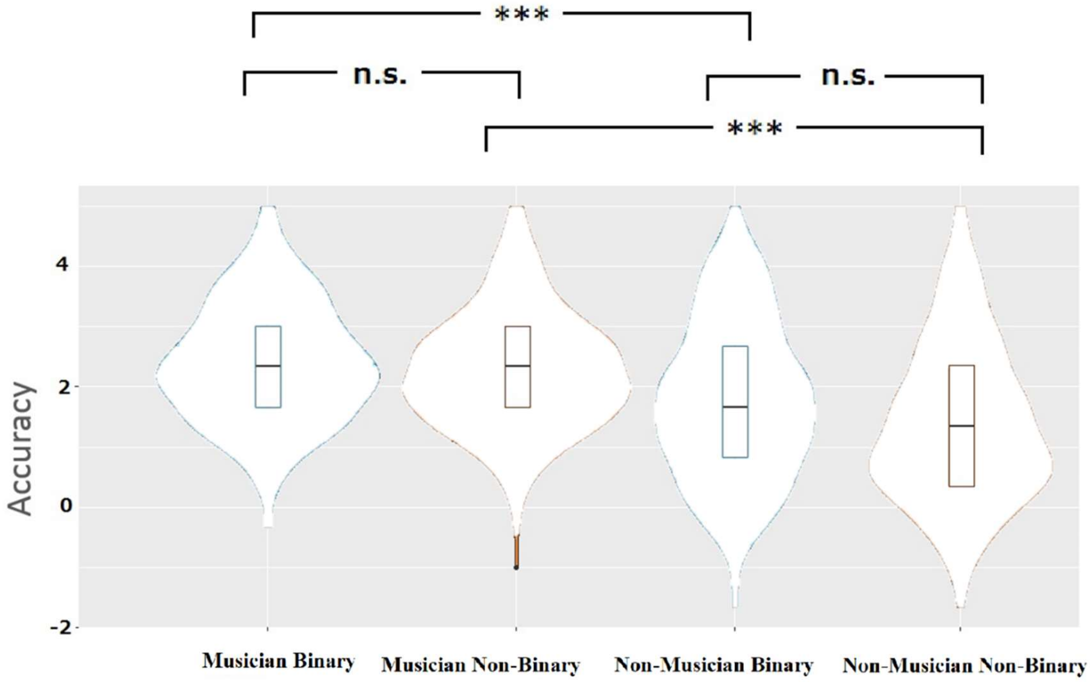


Figure 7.3: Violin plots of accuracy, shown separately for an interaction between rhythm type and musical experience. Data from linguistic stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.3 and Table 7.2 show the interaction between musical experience and rhythm type. There was no difference between the accuracy of binary rhythm in musicians (mean accuracy = 2.392, SE = 0.058) and non-musicians (mean accuracy = 2.27, SE = 0.055). Similarly, there was no statistical difference between the accuracy of non-binary rhythm in musicians (mean accuracy 1.778, SE = 0.076) and non-musicians (mean accuracy = 1.398, SE = 0.073). Figure 7.3 shows that the accuracies of both musicians and non-musicians were higher in detecting rhythm violations in binary than non-binary rhythm.

7.3.2 Rating

Looking at Table 7.3, the best model output of the LMM for the z-scoresd rating showed the effects of language and interaction among violation type and musical experience. However, there was no effect of musical experience, and there were no interactions among language and rhythm type, or between language and violation type.

Table 7.3: Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Musical experience*Rhythm type	1, 1537.06	1.1720	n.s.
Musical experience*Language	1, 117.51	2.0282	n.s.
Rhythm type	1, 1537.90	2.1747	n.s.
Language	**1, 118.36**	**14.786**	**0.00019**
Musical experience*Violation type	**3, 1539.42**	**6.508**	**0.00022**

Table 7.4: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
Musician clash vs. Non-musician clash	0.121	320	1.562	n.s.
Musician clash vs. Musician lapse	**0.727**	**1540**	**13.297**	**<0.0001**
Musician clash vs. Musician sp	-0.1138	1538	-1.557	n.s.
Non-musician clash vs. Non-musician lapse	**0.631**	**1537**	**12.233**	**<0.0001**
Non-musician clash vs. Non-musician sp	**-0.5246**	**1537**	**-7.61**	**<0.0001**
Non-musician clash vs. Non-musician lapse	**0.631**	**1537**	**12.233**	**<0.0001**
Non-musician clash vs. Non-musician sp	**-0.524**	**1537**	**-7.61**	**<0.0001**
Musician lapse vs. Non-musician lapse	0.025	215	0.359	n.s.
Musician lapse vs. Musician sp	**-0.841**	**1538**	**-12.239**	**<0.0001**
Non-musician lapse vs. Non-musician sp	**-1.1561**	**1537**	**-17.826**	**<0.0001**
Musician sp vs. Non-musician sp	-0.289	659	-2.994	n.s.

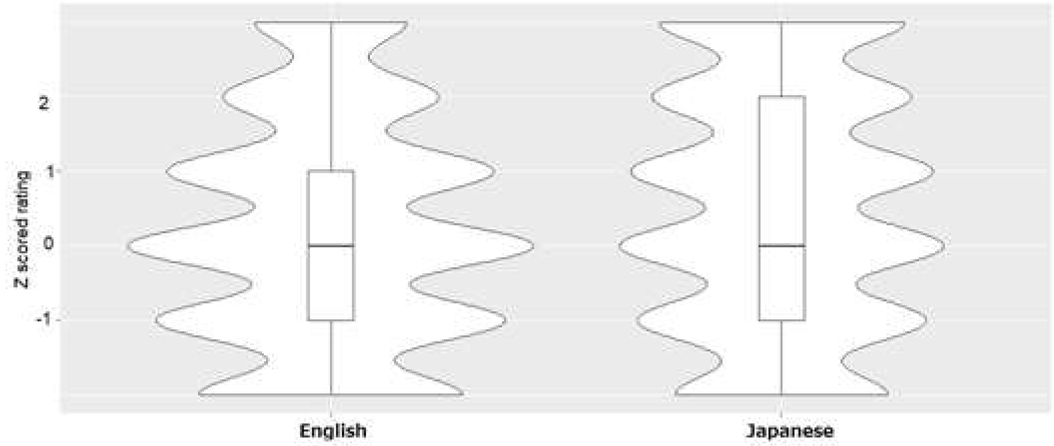


Figure 7.4: Violin plots of z-scores, shown separately for each language. Data from linguistic stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

It is apparent from Figure 7.4 that there was a statistical difference between the z-scoresd rating of English speakers (mean rating = 0.145, SE = 0.032) and Japanese speakers (mean rating = 0.362, SE = 0.034). Japanese speakers rated rhythmic violations higher (more severely) than English speakers.

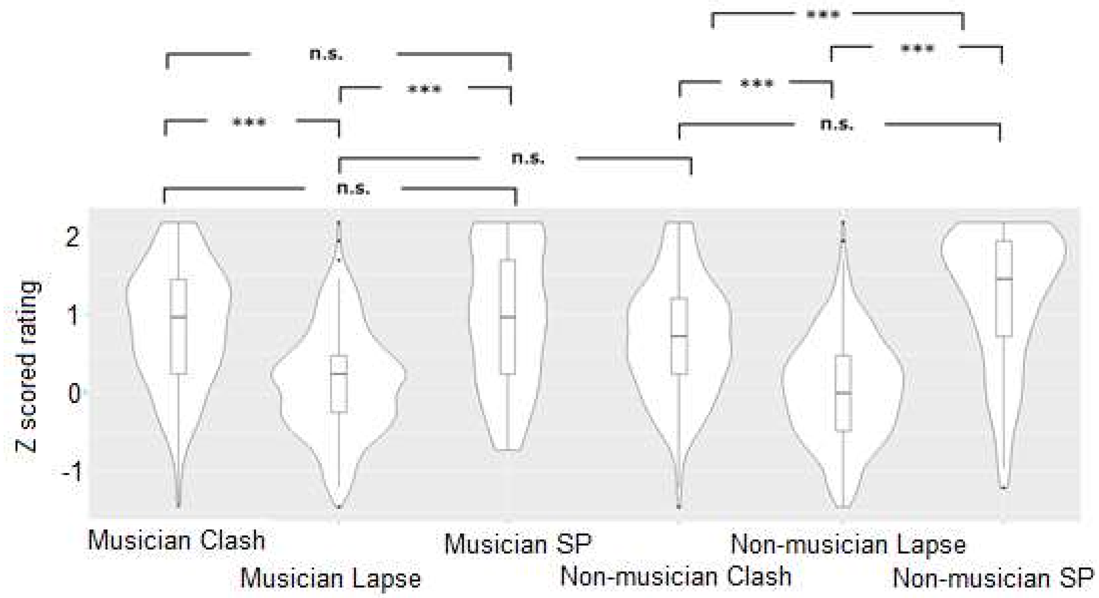


Figure 7.5: Violin plots of z-scores, shown separately for interaction between musical experience and violation type. Data from linguistic stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.5 and Table 7.4 show that there was no statistical difference between musicians and non-musicians in z-scores on each violation type (musicians: clash = 0.842 [SE = 0.031] lapse = 0.116 [SE = 0.033] SP = 0.958 [SE = 0.029], non-musicians: clash = 0.698 [SE = 0.034] lapse = 0.067 [SE = 0.029] SP = 1.220 [SE = 0.031]). The only difference was that there was no statistical difference between z-scores on clash and SP by musicians, while the rating for SP by non-musicians was higher than their rating for clash.

7.3.3 Summary of the results of linguistic stimuli

Regarding accuracy, musical experience did not affect the perception of binary and non-binary rhythm. Similarly, musical experience did not affect the accuracy of violation type, which would mean that musicians are not sensitive to clash, lapse, binary, or non-binary rhythm simply because they are more familiar to such rhythms in musical works. Therefore, this suggests that results showing Japanese accuracy to be higher than English accuracy are due to linguistic experience.

In considering rating, musical experience did not affect the perception or tolerance of binary and non-binary rhythm, or clash, lapse and SP rhythms. Like the results for accuracies, ratings by Japanese speakers were higher than those by the English group, and this would be due to linguistic experience.

7.4 Musical stimuli

7.4.1 Accuracy

As can be seen in Table 7.5, the best model output of the LMM for accuracy showed the effects of language and violation type, and interaction between rhythm type and musical experience. Hoever, the model did not show the effect of musical experience and interaction among language and rhythm type or between language and violation type.

Table 7.5: Statistical summary of fixed effects

Variable	df	F	p
Musical experience*Language	1, 117.22	0.101	n.s.
Musical experience*Violation type	2, 1305.89	1.265	n.s.
Violation type	**2, 1307.9**	**405.587**	**<0.0001**
Language	**1, 118.23**	**11.131**	**0.001**
Rhythm type*Musical experience	**1, 1308**	**12.86**	**0.0003**

Table 7.6: Statistical summary of LSMEANS differences (post hoc)

Variables	Estimate (difference between variables)	df	t	p
Musician binary vs. musician non-binary	**0.464**	**1306**	**7.074**	**<0.0001**
Musician binary vs. non-musician binary	-0.059	167	-0.479	n.s.
Musician non-binary vs. non-musician non-binary	**0.268**	**169**	**2.164**	**<0.0001**
Non-musician binary vs. non-musician non-binary	**0.7919**	**1307**	**12.555**	**<0.0001**

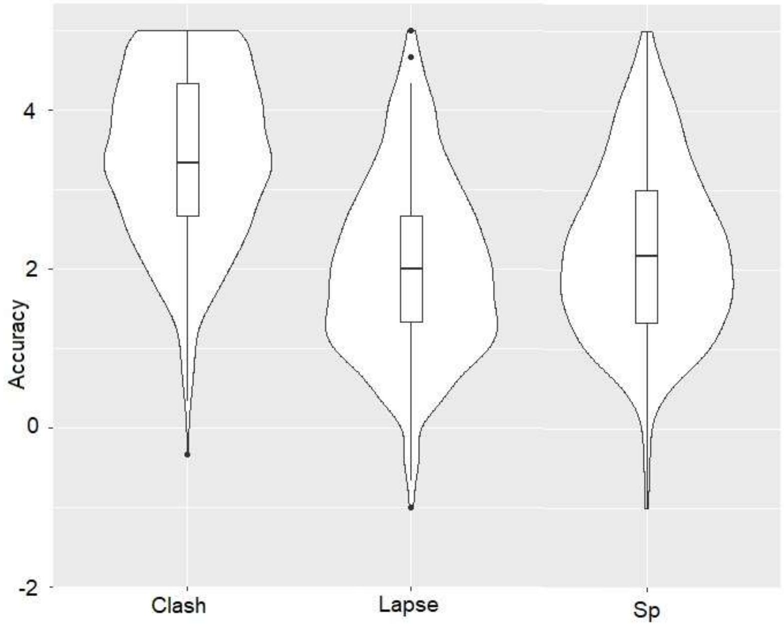


Figure 7.6: Violin plots of accuracy, shown separately for each violation type. Data from musical stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.6 shows that listeners were more accurate in detecting clashes (mean accuracy = 3.390, SE = 0.048) than lapses (mean accuracy = 1.932, SE = 0.042). It also shows that the accuracy of SP was lower than clash (mean accuracy = 1.872, SE = 0.041).

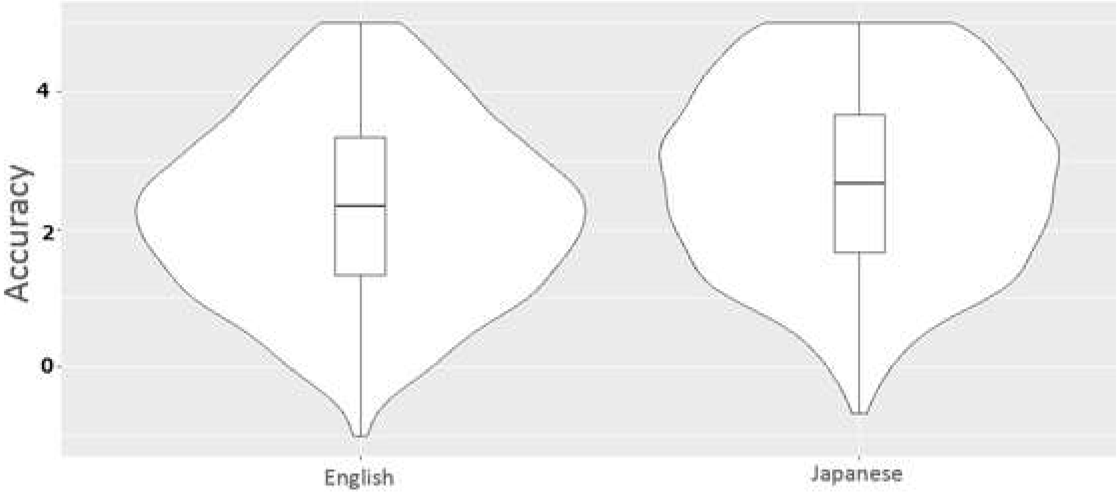


Figure 7.7: Violin plots of accuracy, shown separately for each language. Data from musical stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.7 demonstrates that the accuracy of Japanese speakers (mean accuracy = 2.752, SE = 0.051) was higher than that of English speakers (mean accuracy = 2.25, SE = 0.053). As there was no interaction among rhythm type (binary or non-binary) and native language, Japanese accuracy of binary and non-binary rhythms were higher than those of English speakers.

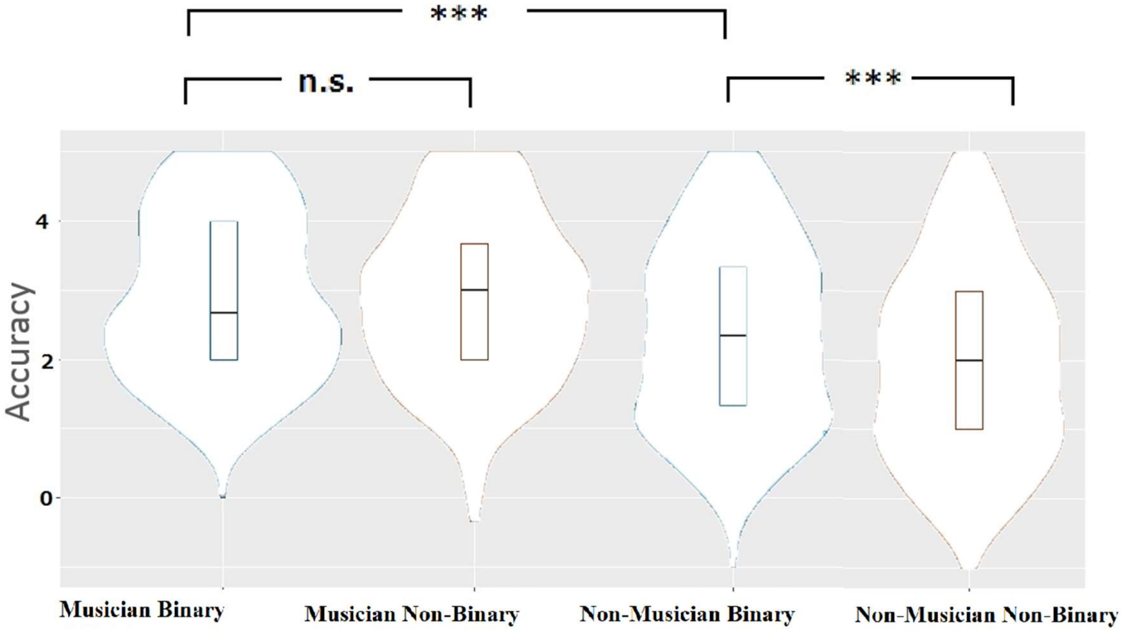


Figure 7.8: Violin plots of accuracy, shown separately for an interaction between rhythm type and musical experience. Data from musical stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.8 and Table 7.6 display the interaction between musical experience and rhythm type. Looking at the difference in accuracy between musicians and non-musicians, there was a statistical difference between the non-binary accuracy of musicians (mean accuracy = 2.358, SE = 0.077) and non-musician's (mean accuracy = 1.954, SE = 0.077), while there was no difference between the binary rhythm accuracy of musicians (mean accuracy = 2.899, SE = 0.068) and non-musicians (mean accuracy = 2.867, SE = 0.065).

7.4.2 Rating

From Table 7.7, we can see that the best model output of the LMM for rating revealed an effect of language and violation type and interaction between rhythm type and musical experience.

Table 7.7: Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Musical experience*Language	1, 117.3	0.021	n.s.
Musical experience* Violation type	3, 1546.62	1.795	n.s.
Violation type	**3, 1549.61**	**931.3529**	**<0.0001**
Language	**1, 118.28**	**9.0974**	**0.0031**
Musical experience*Rhythm type	**1, 1550.42**	**8.696**	**0.003**

Table 7.8: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
Musician binary vs. Non-musician binary	-0.1209	183	-1.646	n.s.
Musician binary vs. Musician non-binary	**0.13**	**1549**	**3.07**	**0.011**
Non-musician binary vs. Non-musician non-binary	**0.303**	**1547**	**7.535**	**<0.0001**

Musician non-binary vs. Non-musician non-binary	0.051	182	0.705	n.s.

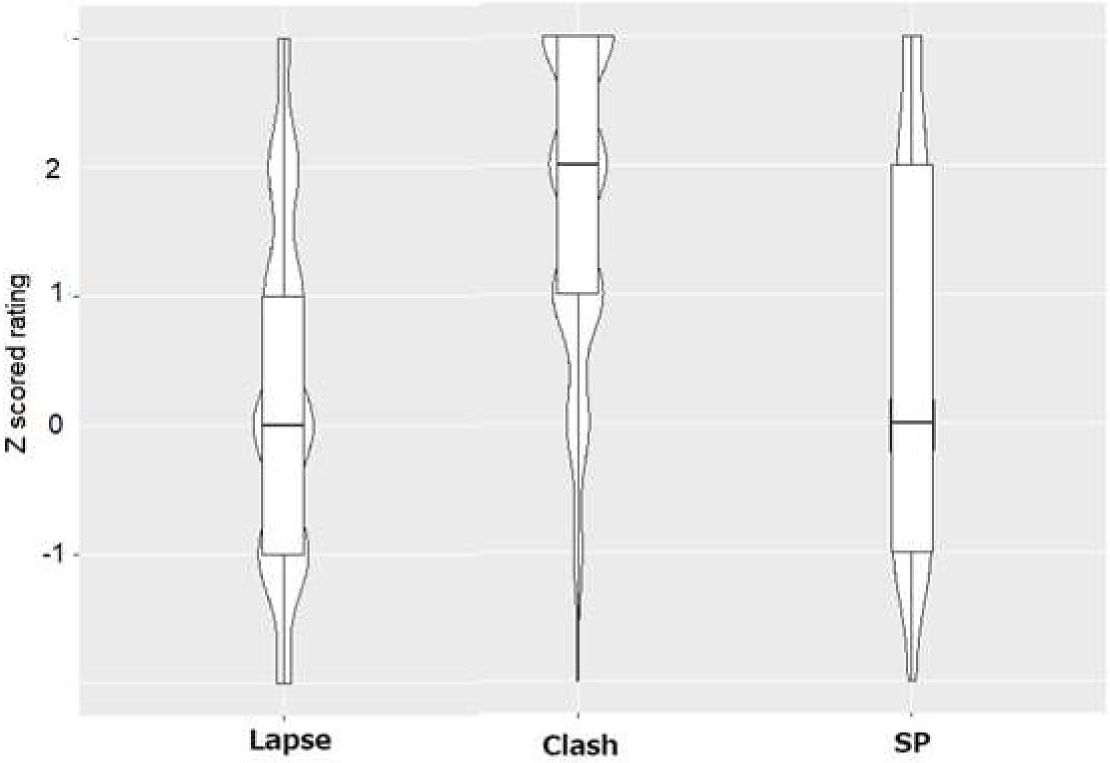


Figure 7.9: Violin plots of z-scores, shown separately for each violation type. Data from musical stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Looking at Figure 7.9, the rating of clash (mean = 1.28, SE = 0.06) was higher than other violation types, while the rating of lapse (mean = 0.542, SE = 0.611) was lowest. The rating of SP (mean = 0.342, SE = 0.064) was second highest.

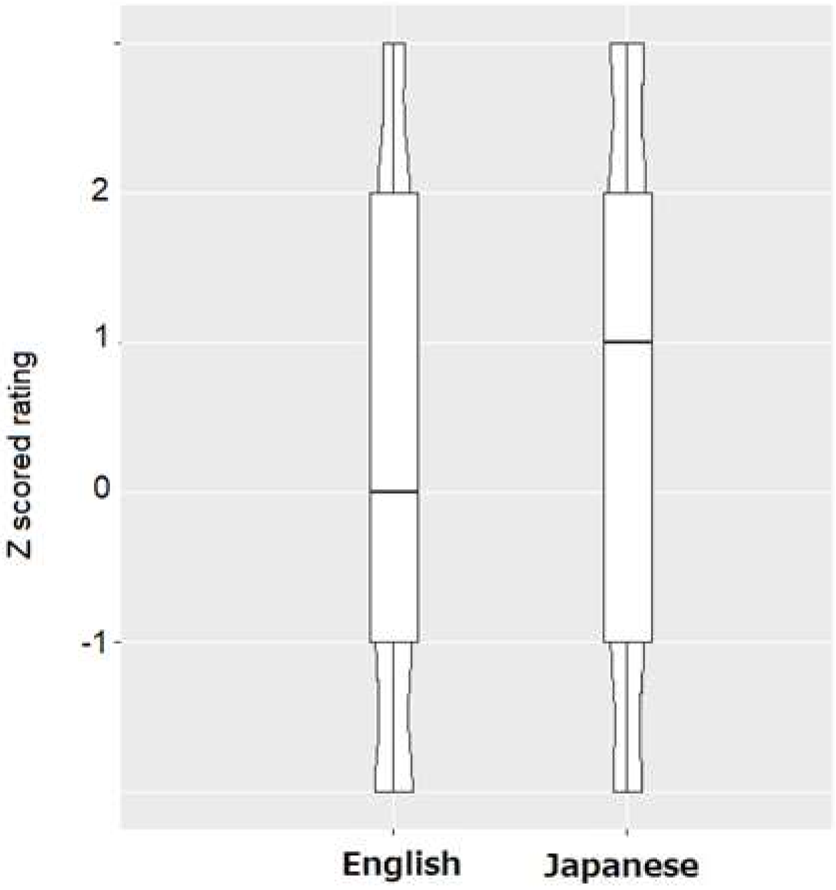


Figure 7.10: Violin plots of z-scores, shown separately for each language. Data from musical stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.10 shows that there was a statistical difference between English (mean = 0.401, SE = 0.064) and Japanese (mean = 0.723, SE = 0.064) speakers. The ratings of Japanese speakers were higher than those of English speakers. This result does not differ from the results of ratings in the previous chapter and sections.

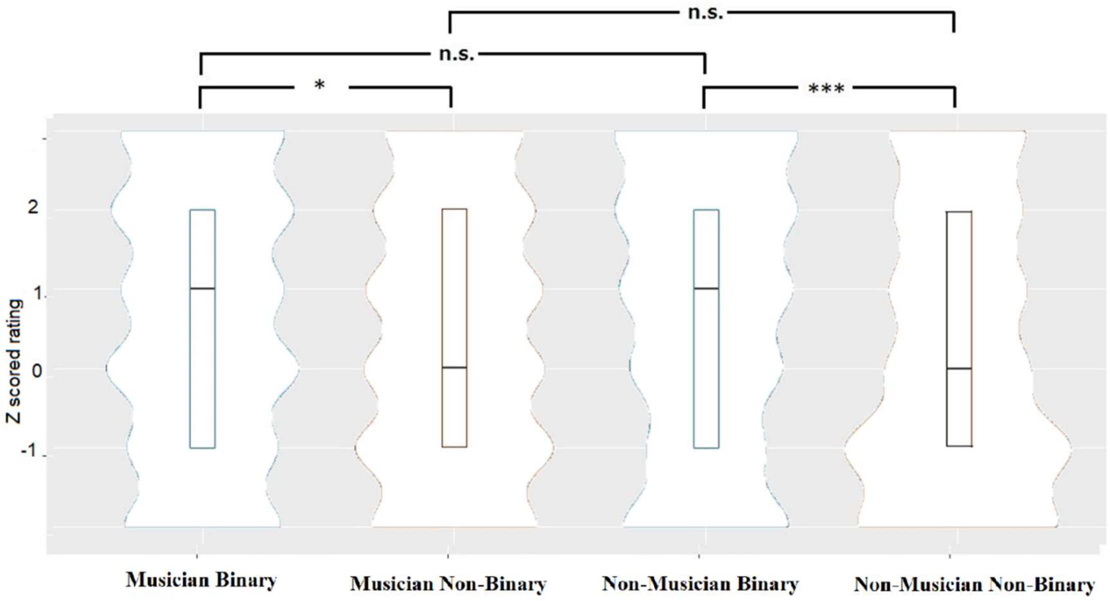


Figure 7.11: Violin plots of z-scores, shown separately for each violation type. Data from musical stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

As shown in Figure 7.11 and Table 7.8, there was a difference between the rating of binary rhythm by musicians (mean rating = 0.355, SE = 0.053) and non-binary rhythm (mean rating = 0.226, SE = 0.051) and between the rating of binary rhythm by non-musicians (mean rating = 0.439, SE = 0.049) and non-binary rhythm (mean rating = 0.137, SE = 0.047). Moreover, z-scoresd ratings for binary rhythm by both musicians and non-musicians were higher than those for non-binary rhythm (rating of binary rhythm was higher than non-binary rhythm). However, there was no difference between musician and non-musician.

7.4.3 Summary of the results of musical stimuli

Looking at accuracy, musical experience affected the accuracy of non-binary rhythm, while it did not affect the accuracy of binary rhythm. The only difference compared to the previous section on linguistic stimuli was that accuracy on non-binary rhythm for musicians was higher than that of non-musicians. This would suggest that musical experience enhances the accuracy of non-binary rhythm in musical phrases.

In considering the ratings, there was no difference between musicians and non-musicians, regardless of violation type and rhythm type. Although Japanese ratings were higher than English speakers, this would not be due to musical experience but rather to linguistic experience, as musical experience did not affect the ratings in musical stimuli.

7.5 Tonal stimuli

7.5.1 Accuracy

There was no difference between the accuracies of long tonal stimuli and short tonal stimuli (linear mixed model, t = -0.535, df = 2336; p = n.s.); thus, the two types of pure tone stimuli were treated as a tonal stimuli group and further analysed.

It can be seen from Table 7.9 that there were effects of language and interaction between rhythm type and musical experience and between musical experience and violation type.

Table 7.9: Statistical summary of fixed effects

Variable	**Df**	**F**	**p**
Musical experience*Language	1, 117.44	0.369	n.s.
Language	**1, 118.47**	**14.503**	**0.0002**
Musical experience*Violation type	**2, 2724**	**13.195**	**<0.0001**
Musical experience*Rhythm type	**1, 2924.99**	**3.98**	**0.046**

Table 7.10: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
Musician clash vs. Non-musician clash	0.181	202	1.396	n.s.
Musician clash vs. Musician lapse	**1.433**	**1306**	**19.578**	**<0.0001**
Musician clash vs. Musician sp	**1.266**	**1306**	**12.909**	**<0.0001**
Non-musician clash vs. Non-musician lapse	**1.409**	**1307**	**20.027**	**<0.0001**
Non-musician clash vs. Non-musician sp	**1.06**	**1306**	**11.345**	**<0.0001**
Musician lapse vs. Non-musician lapse	0.156	160	1.281	n.s.
Musician lapse vs. Musician sp	-0.167	1307	-1.801	n.s.
Non-musician lapse vs. Non-musician sp	**-0.348**	**1307**	**-3.927**	**0.001**
Musician sp vs. Non-musician sp	-0.243	359	-0.16	n.s.
Musician binary vs. Non-musician binary	-0.0164	173	-0.167	n.s.
Musician binary vs. Musician non-binary	**0.582**	**2725**	**10.751**	**<0.0001**
Non-musician binary vs. Non-musician non-binary	**0.732**	**2726**	**14.074**	**<0.0001**
Musician non-binary vs. Non-musician non-binary	0.133	175	1.357	n.s.

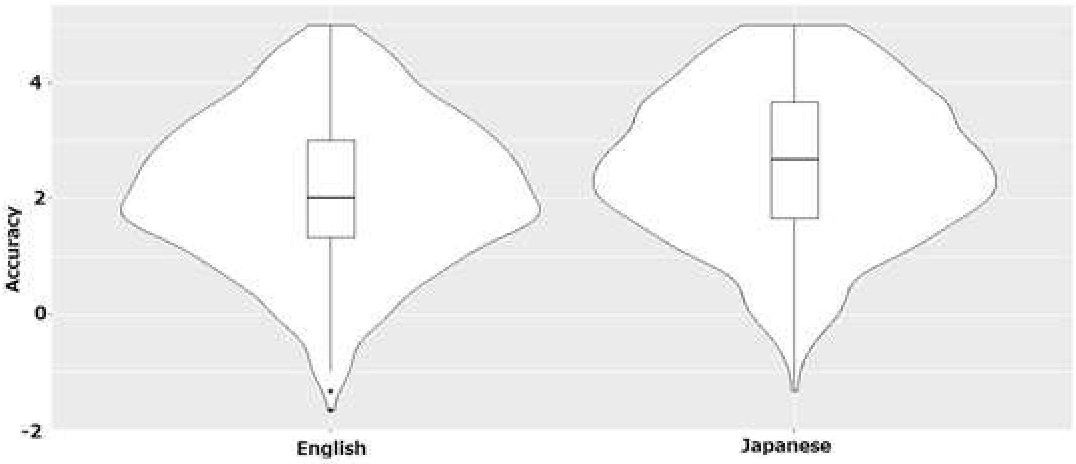


Figure 7.12: Violin plots of accuracy, shown separately for each language. Data from tonal stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

From Figure 7.12, it is apparent that the accuracy of Japanese speakers (Japanese accuracy = 2.547, SE = 0.036) was higher than that of English speakers (mean accuracy = 2.16, SE = 0.088). This result is identical to all previous results on the differences between the English and Japanese groups.

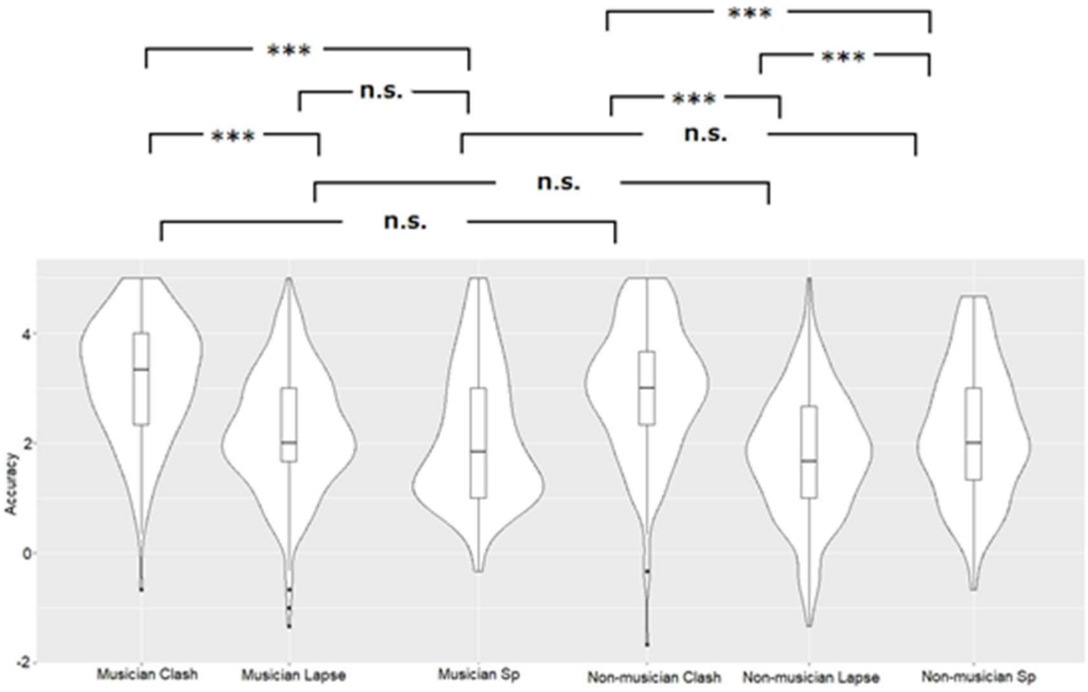


Figure 7.13: Violin plots of accuracy, shown separately for an interaction between musical experience and violation type. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.13 and Table 7.10 show that there was no statistical difference between musicians and non-musicians in accuracies on each violation type (musicians: clash = 3.175 [SE = 0.079] lapse = 2.125 [SE = 0.071] SP = 2.004 [SE = 0.081], non-musicians: clash = 3.055 [SE = 0.069] lapse = 1.808 [SE = 0.071] SP = 2.196 [SE = 0.075]). The only difference was that there was no statistical difference between accuracy on lapse and SP by musicians, while accuracy on SP by non-musicians was higher than their accuracy on lapse.

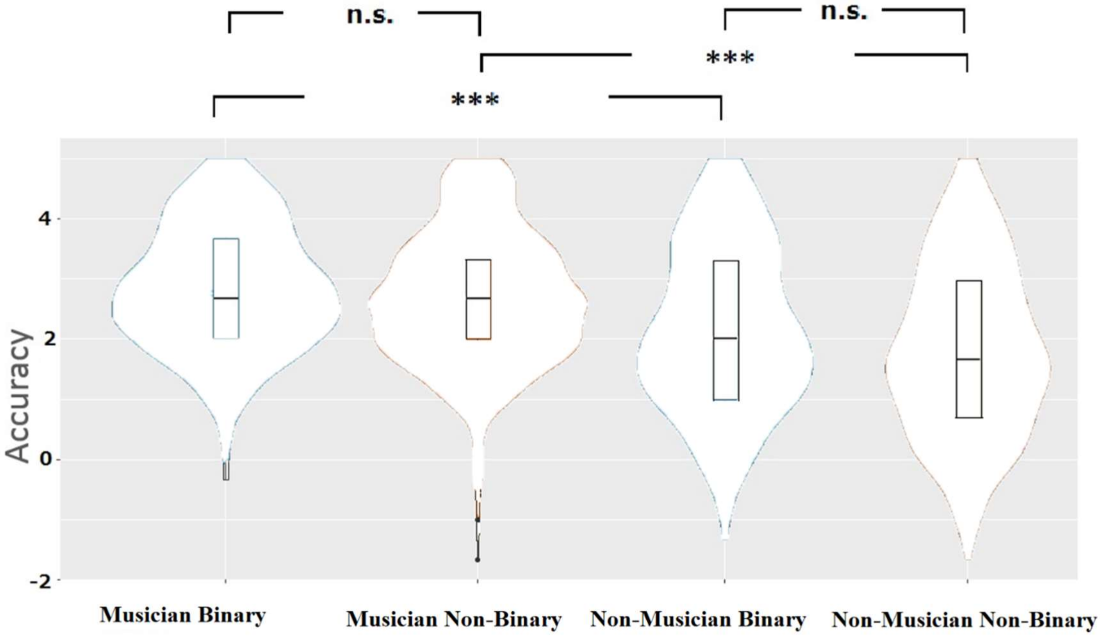


Figure 7.14: Violin plots of accuracy, shown separately for an interaction between musical experience and rhythm type. Data from Russian speakers was excluded. Data from musicians was included.

As can be seen from Figure 7.14, the accuracy of binary rhythm of musicians (mean accuracy = 2.84, SE = 0.042) was not different from that of non-musicians (mean accuracy = 2.675, SE = 0.044), and accuracy of non-binary rhythm of musicians (mean accuracy = 2.161, SE = 0.056) was also identical to that of non-musicians (mean accuracy = 1.802, SE = 0.056). The accuracy of binary rhythm was higher than non-binary rhythm in both musician and non-musician groups. These results are identical to the result of linguistic stimuli.

7.5.2 Rating

There was no difference between the accuracy of long pure tone stimuli and short pure tone stimuli (linear mixed model, t = -0.19, df = 1.825e+04; p = n.s.); thus, the two types of pure tone stimuli were treated as a pure tone stimuli group and further analysed.

Table 7.11 shows the main effects of language. Also, there was an interaction among rhythm type and musical experience and among violation type and musical experience.

Table 7.11: Statistical summary of fixed effects

Variable	**df**	**F**	**p**
Musical experience*Language	1, 118	1.5133	n.s.
Language	**1, 119**	**8.281**	**0.0047**
Musical experience*Violation type	**3, 3208.3**	**12.2271**	**<0.0001**
Musical experience*Rhythm type	**1, 3208.2**	**11546**	**0.0006**

Table 7.12: Statistical summary of LSMEANS differences (post hoc)

Variables	**Estimate (difference between variables)**	**df**	**t**	**p**
Musician binary vs. Non-musician binary	**-0.1495**	**186**	**-2.661**	**0.041**
Musician binary vs. Musician non-binary	**0.143**	**3209**	**4.308**	**0.0001**
Non-musician binary vs. Non-musician non-binary	**0.2977**	**3207**	**9.547**	**<0.0001**
Musician non-binary vs. Non-musician non-binary	0.0052	185	0.093	n.s.
Musician clash vs. Non-musician clash	-0.0195	270	-0.314	n.s.
Musician clash vs. Musician lapse	**0.821**	**3209**	**20425**	**<0.0001**
Musician clash vs. Musician sp	**0.873**	**3208**	**16.168**	**<0.0001**
Non-musician clash vs. Non-musician sp	**0.632**	**3208**	**12.503**	**<0.0001**
Musician lapse vs. Non-musician lapse	0.1255	191	2.218	n.s.
Musician lapse vs. Musician sp	0.052	3209	1.027	n.s.
Non-musician lapse vs. Non-musician sp	**-0.333**	**3208**	**-6.987**	**<0.0001**
Musician sp vs. Non-musician sp	**-0.2601**	**562**	**-3.462**	**0.0133**

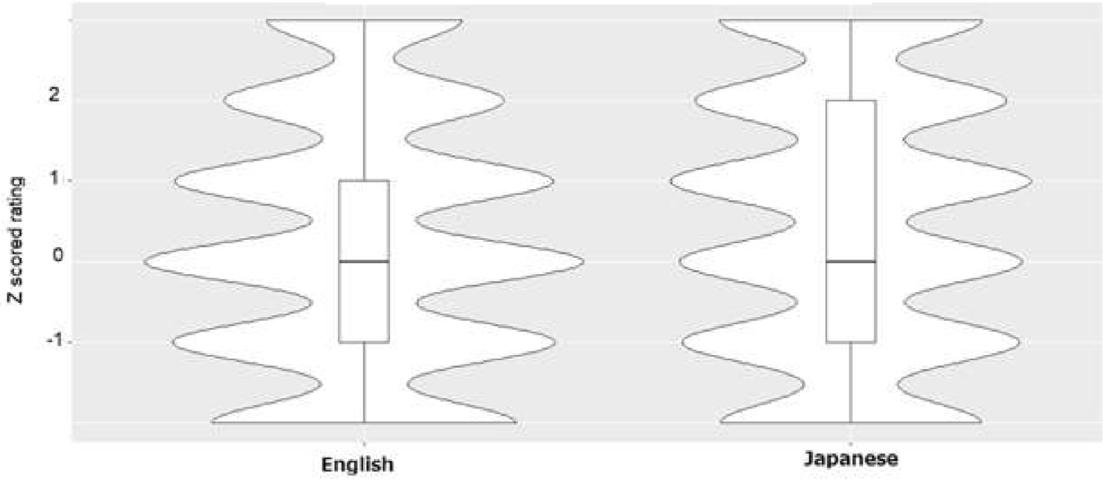


Figure 7.15: Violin plots of z-scores, shown separately for each language. Data from tonal stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

It is apparent from Figure 7.15 that z-scoresd rating by Japanese speakers (mean rating = 0.342, SE = 0.025) was higher than that of English speakers (mean rating = 0.196, SE = 0.024).

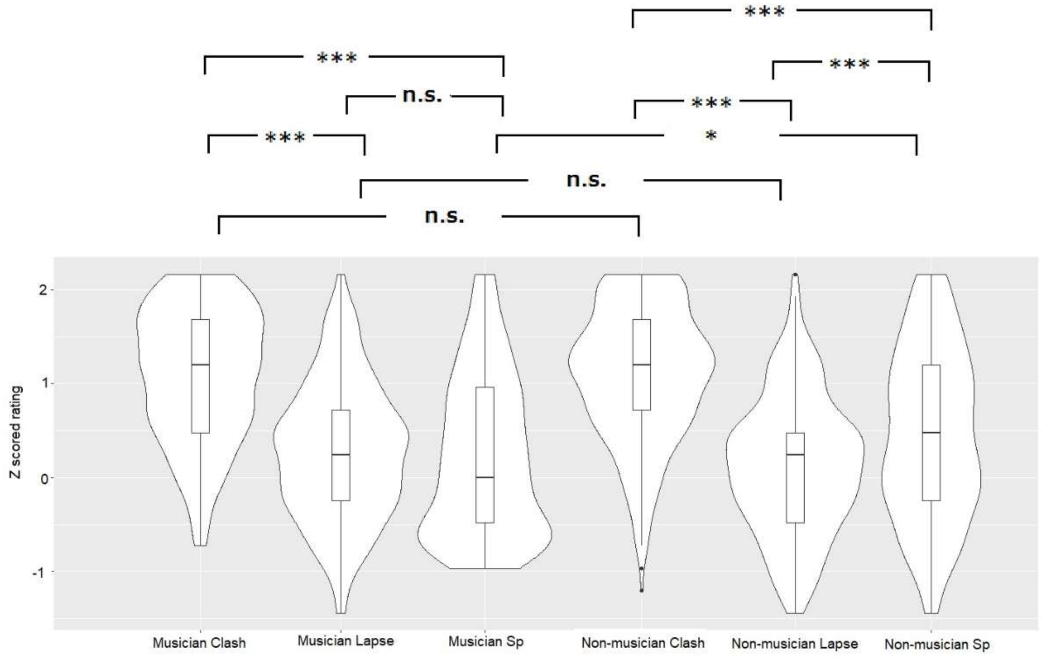


Figure 7.16: Violin plots of z-scores, shown separately for interaction between musical experience and violation type. Data from tonal stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.16 and Table 7.12 show that there was no statistical difference between musicians and non-musicians in accuracies on each violation type, except for SP (musicians: clash = 1.09 [SE = 0.031] lapse = 0.269 [SE = 0.038] SP = 0.214 [SE = 0.031], non-musicians: clash = 1.092 [SE = 0.025] lapse = 0.124 [SE = 0.041] SP = 0.459 [SE = 0.029]). Rating for SP by musicians was lower than that of non-musicians.

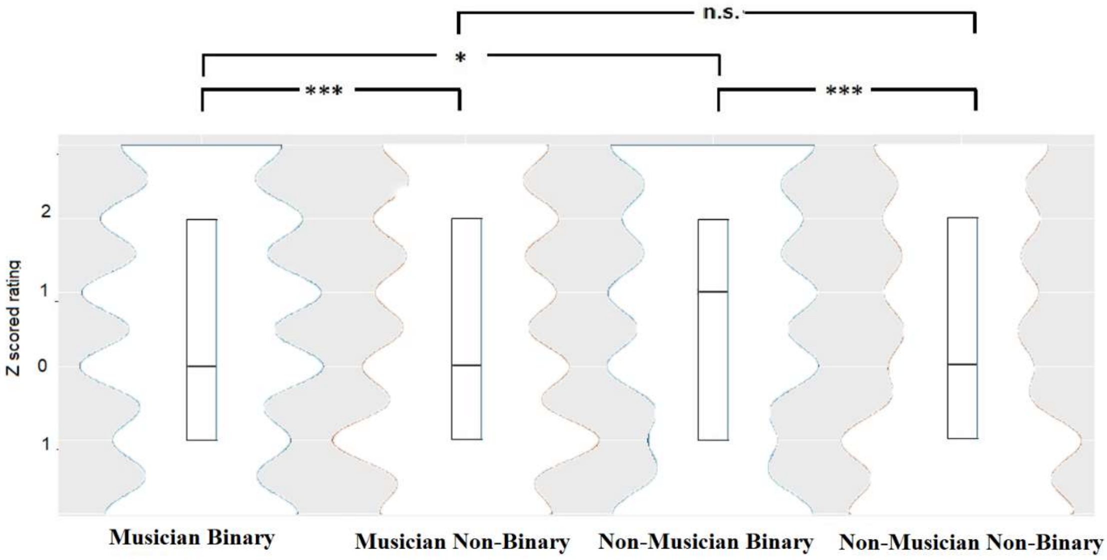


Figure 7.17: Violin plots of z-scores, shown separately for interaction between musical experience and rhythm type. Data from tonal stimuli only. Data from Russian speakers was excluded. Data from musicians was included.

Figure 7.17 and Table 7.12 show that there was a difference between rating of non-binary rhythm by musicians (mean rating = 0.212, SE = 0.035) and non-musicians (mean rating = 0.119, SE = 0.033). However, there was also a difference between rating of binary rhythm by musicians (mean rating = 0.34, SE = 0.035) and non-musicians (mean rating = 0.423, SE = 0.032). With respect to the binary rhythm, and contrary to the intuitive view, rating by non-musicians was actually higher than that of musicians. Ratings of binary rhythm by both musicians and non-musicians were statistically higher than those of non-binary rhythms.

7.5.3 Summary of the results of pure-tone stimuli

With respect to accuracy, there was no difference between musician and non-musician groups in violation type (clash, lapse, and SP) and rhythm type (binary and non-binary rhythms) akin to the results of the linguistic experiment.

Looking at ratings, there was no difference between musicians and non-musicians in tonal stimuli, except for SP and binary rhythm. Rating for SP by musicians was lower than that of non-musicians, whereas rating for binary rhythm by musicians was lower than that of non-musicians.

7.6 Discussion

The main purpose of this chapter was to examine if, and how, musical experience affects rhythm perception. While some studies, such as Kishon-Rabin, Amir, Vexler, and Zaltz (2001), show that musical experience boosts pitch perception, the role of musical experience on rhythm perception has remained unclear. With respect to the accuracy with which violations are detected in stimuli with binary or non-binary rhythm, musicians detected violations in non-binary rhythm better than non-musicians, but this applied only to musical stimuli. This musical influence on accuracy in non-binary rhythm is identical to the results of Hannon and Trehub (2005), in which participants with musical experience performed better in non-binary rhythm musical stimuli, but not in binary stimuli. Drake (1993), who showed musicians perform better than non-musicians in irregular rhythm trials, supports the result. A possible reason why musicians only perform better in musical stimuli is that they are more familiar with musical sounds than non-musicians, while there is no difference of familiarity between musicians and non-musicians in linguistic and tonal stimuli.

For interaction of ratings between musical experience and rhythm type (binary or non-binary rhythm), there was no difference between musician and non-musician groups, which would suggest that musical experience does not affect the ratings for binary and non-binary rhythm. Moreover, in musical stimuli, musicians tended to rate the difference between binary and non-binary rhythm as small, compared with non-musicians. This would mean that musicians tend to judge rhythmic difference on the same perspective as they judge the rhythmic violation in binary and non-binary rhythms. For the ratings for SP by musicians in tonal stimuli, the rating was lower than that of non-musicians. This difference may be due to the tonal stimuli being simpler than other stimuli; thus, musicians could detect that the SP is, as the name suggests, structure-preserving.

Chapter 8: General discussion

8.1 Introduction

The experiments in this thesis showed perceptual differences between English, Japanese, and Russian speakers. It was clear that the accuracies of Japanese speakers were higher than English speakers in irregular rhythm. This difference is related to the large differences in speech rhythm between English and Japanese. The results of Russian speakers were identical to English speakers. Additionally, musicians performed better than non-musicians in non-binary trials, but the influence was limited to musical stimuli. The results of the ratings were similar to those for accuracies; in other words, Japanese rating was higher than that of the English group, while Russian rating was not statistically different from the English speakers. The influence of musical experience on the ratings was smaller than linguistic influence, in the sense that musical experience did not affect the ratings. The following hypotheses were tested in this study:

Hypothesis 1: All participants, independently of language, will find clashes worse than lapses: they will rate stimuli with clashes as notably different and will be more accurate at detecting differences when stimuli have clashes, compared to those with lapses.

This hypothesis was proven to be true in all the results in the previous two chapters, regardless of linguistic and musical backgrounds. In Section 2.2.2 and 2.3.2, it was shown that lapses were better tolerated than clashes.

Hypothesis 2: It was hypothesised that the linguistic and musical experience influence the rhythmic perception based on findings by Hannon and Trehub (2005), who showed that people who were familiar with non-binary rhythm are better at detecting violations to non-binary rhythmic structure compared to people who are unfamiliar with that kind of rhythm.

This hypothesis was supported by the results: Japanese participants were more capable of detecting rhythm irregularities in stimuli with non-binary rhythm, compared to English and Russian participants. However, the results showed that Japanese participants also performed better than the English and Russian groups in binary rhythm trials, even though the latter two groups were more familiar with binary rhythms than the Japanese speakers. This is possibly due to subjective rhythmisation, which allows listeners to anticipate beats. Therefore, English and Russian participants could hear illusional binary beats and filled in lapses. This result is comparable to Grabe and Warren (1995) and Hannon and Trehub (2005), who state that participants perceived binary rhythm even when they listened to non-binary rhythm. Subjective rhythmisation occurred in the results of participants who were unfamiliar with non-binary rhythm.

Hypothesis 3: Looking at Section 2.2.2., lapses frequently occur in Japanese. Considering this, along with Hannon and Trehub's (2005) finding that participants were more sensitive to familiar rhythm than unfamiliar rhythm, Japanese speakers will be more sensitive to lapses than English and Russian speakers.

The results in Chapter 6 showed that there was no interaction between native language and violation type, which means that accuracies of clash by English, Japanese, and Russian speakers were similarly higher than those of lapse. This is supported by the result of hypothesis 1 – that there was no difference in preference of an irregular rhythm (clash and lapse) between the groups of participants.

Hypothesis 4: Linguistic experience will affect how each group treats linguistic stimuli and musical experience will affect how they treat musical stimuli. As musicians are more familiar with musical rhythm than non-musicians, musicians will be more sensitive to musical stimuli. This type of difference between musicians and non-musicians will be lessened in the experiments with tonal and linguistic stimuli.

This hypothesis was partially supported; however, the influence of musical experience was limited to musical stimuli with non-binary rhythm. Experience with speech rhythm, on the other hand, was significantly more far reaching, influencing accuracies within all the stimulus types. This is evident, for instance, in the accuracies of the Japanese speakers, which were higher than those of English speakers in all modalities, while the accuracies of Russian speakers were identical to those of the English speakers. This difference between the results of linguistic and musical experience comparisons suggests that the influence of linguistic experience is greater than that of musical experience.

Hypothesis 5: Considering Drake (1993), who demonstrates that musicians performed better in reproducing complicated rhythm than non-musicians, and that there was no significant difference between musicians and non-musicians in reproducing simple rhythm, we can hypothesise that musicians will perceive non-binary rhythm better than non-musicians, while there will be no difference between musicians and non-musicians in perceiving binary rhythm.

The results in Chapter 7 showed that, as hypothesised, the accuracy of musicians in non-binary rhythm was higher than that of non-musicians. However, this difference applies only to trials involving non-binary rhythms (cf. Drake, 1993). A possible explanation is that binary rhythms were sufficiently simple that all participants performed well.

In addition to the hypotheses above, this thesis examined in some detail the nature of Japanese speech rhythm. The results of the corpus study showed that Japanese rhythm is trochaic, but that the trochaic rhythm in Japanese is subtle, as shown in Chapter 4.

8.2 Accuracy

In most experiments, English and Russian speakers, whose linguistic rhythms are similar, had similar rhythmic perception, while Japanese speakers, who are supposed to be familiar with non-binary rhythm due to both musical and linguistic experiences, more accurately detected rhythmic differences between stimuli than the English and Russian speakers (see Section 6.6).

In some cases, it was unclear whether the results of the linguistic experience comparison groups were influenced by native language or music education. Considering the result that accuracy of the musicians group was higher only in the non-binary rhythm of the musical stimuli, the differences of accuracies between English and Japanese speakers in a binary rhythm would at least be influenced, if not by a difference in music education, then by native language. It might appear that the English accuracy of non-binary rhythm in linguistic stimuli was lower than Russian speakers because of a difference in musical experience – as musical experiences between English and Russian speakers are different (see Section 3.5) – and because non-binary rhythm is tolerated in Russian music. However, considering the result that musical experience increased only the accuracy in the non-binary rhythm of musical stimuli, the difference of musical cultural background between English and Russian speakers does not seem to affect the accuracy of non-binary rhythm in linguistic stimuli, because the English and Russian accuracies of non-binary rhythm in musical stimuli were statistically identical.

Therefore, the reason that Russian speakers achieved a higher level of accuracy of non-binary rhythm in linguistic stimuli than English speakers will not be because of the difference in musical experience, but possibly the difference in native language (i.e. tolerance of lapse in Russian, given that another cultural difference does not affect rhythm perception).

Considering the results based on the difference of cues, a possible explanation for the difference in the accuracies in linguistic stimulus between English and Russian speakers would be that the least important cue to detect accent in English is pitch (Kochanski, Grabe, Coleman, & Rosner, 2005), which played the role of accent in the linguistic stimuli.

Except for the accuracy of non-binary rhythm in linguistic stimuli, all English accuracies were statistically consistent with those of the Russian speakers. This would suggest that a difference of musical education between countries does not affect rhythm perception. In conclusion, although there was a difference between musicians and non-musicians in the accuracy of non-binary rhythm in musical stimuli, music education in each country and cultural musical rhythm does not seem to significantly affect rhythmic accuracy.

8.3 Rating

With respect to the relationship between Japanese and Russian speakers, ratings of lapse and SP tended to be statistically similar. These can be explained by focusing on the tolerance for lapses in these languages and by the mora-counting of SP by Japanese speakers. In most cases, Japanese ratings were higher than those of the Russian group; however, there was a tendency for Japanese speakers to rate SP lower, based on moraic perception. As a result, Japanese and Russian ratings of SP were statistically identical.

Moreover, Japanese ratings of clash were consistently higher than English and Russian speakers, except for in pure-tone stimuli. A possible reason for this severe judge of clash by the Japanese group would be that clash is impossible in Japanese, as mentioned in Chapter 1, while clash is possible in the English and Russian languages.

8.4 Conclusion

With respect to the corpus-based study for Japanese rhythm, contrary to the popular belief that Japanese is a mora-timed language – meaning that the moraic duration of Japanese phrase was supposed to be constant, as discussed in Bloch (1950) and Jinbo (1980) – it was shown that the Japanese rhythm is a subtle trochaic. Similar to stress-timed language and syllable-timed language, experimental studies, such as Kato (1999) and Kato et al. (1997), could not prove that Japanese is a mora-timed language. However, the results of the current study showed that Japanese is a subtle trochaic rhythm. The results also showed the effectiveness of the corpus for the rhythm study and the role of pitch accent as prominence. Haraguchi (1991) and Tajima (1998) both suggested that Japanese pitch accent works similarly to stress accent, in the sense that both stress and pitch accents play a role as prominence in linguistic rhythm – although Haraguchi (1991) and Tajima (1998) did not show this through data or experiments. The results reported in Chapter 4 showed that pitch accent can be prominent, making it a candidate for the foundation of speech rhythm in Japanese. As the pitch accent was located on odd-numbered morae to make the rhythm trochaic, the Japanese rhythm can be said to be trochaic.

It is important to consider the dynamic attending theory proposed by Jones (1976), which mentions that binary rhythm helps us to understand the sound process. This seems to be applicable to language perception, considering Quené and Port (2005), who show that binary rhythm in a language helps us to understand phonemes. The results of the current study may imply that the dynamic attending theory is culture-dependent, considering that Japanese speakers, who are familiar with non-binary rhythm, can communicate with the subtle binary rhythm. In other words, it is possible that English speakers might rely on the binary rhythm to understand linguistic phrases correctly, while Japanese speakers do not. This reliance on the binary rhythm by English speakers could affect the results in the current study.

Looking at rhythm perceptions, Japanese accuracies were higher than English speakers, and Japanese tended to rate the rhythmic difference more severely. With respect to the relationship between Japanese and Russian speakers, there were some similarities in terms of rating and accuracy, possibly due to the fact that Russians tolerate lapse similarly to Japanese (Gouskova & Roon, 2013) and that Japanese eurhythmy is a trochaic alternation with frequent lapse, although the eurhythmy is difficult to achieve due to the lack of the number of accents that cause lapse. However, while it was hypothesised that Japanese speakers will tolerate lapse, which is the irregular rhythm used in the Japanese language, their ratings were similar to those of English and Russian speakers.

It might seem counterintuitive that the accuracies of English and Russian speakers in binary rhythm were lower than those of Japanese speakers, as English and Russian speakers are more familiar with the binary rhythm in their languages than Japanese speakers. Grabe and Warren (1995) mention that English speakers find that clash was avoided by stress shift, even when they hear a phrase with a stress clash: this bias may mean that they don’t perceive all violation types (clash and lapse) as violations. This type of illusional interpretation is found in Hannon and Trehub (2005), who state that English speakers tended to perceive a non-binary rhythm phrase preceded by binary rhythm, as a binary rhythm phrase. This illusional rhythmic perception may be a reason for the low accuracies of English and Russian speakers in binary rhythm conditions.

This illusional binary rhythm might also be caused by subjective rhythmisation, which is said to be a phenomenon where one perceives a sequence of monotonic sounds as regular alternation of prominent and non-prominent sounds; for example, even if a sequence does not have a pattern of prominence and non-prominence, one perceives illusive alternation of prominence and non-prominence. As far as it is known, it is not clear whether one actually perceives non-binary rhythm as binary due to the subjective rhythmisation. However, in Hannon and Trehub (2005), it is shown that North American adults, in the non-binary rhythm condition, tended to perceive the structure-violating stimuli (binary rhythm) as more similar to the original familiarisation stimulus (non-binary rhythm), than the non-violating stimuli (non-binary rhythm). They suggest that this is because participants mistook the original familiarisation stimulus (non-binary rhythm) for a binary rhythm. This false perception may be due to the subjective rhythmisation, which makes listeners perceive a sequence of sounds as binary rhythm (regular alternation of prominent and non-prominent sounds). If so, the results of the current study may be not due to familiarity of the rhythms, but due to subjective rhythmisation, which could work differently depending on the listeners native language.

In all models, it was shown that all speakers favour lapse over clash, as Wasow et al. (2015) suggest (although their participants were limited to English speakers). In the languages tested in the current study, clash was less common than lapse. Similarly, in the musical works considered in this study, clash was not found, while lapse was used in some of the pieces. Looking at the results of Vuust et al. (2009), participants (Finnish speakers) did not tolerate clash than the lapse in musical passages. They also showed that there was no difference between musicians and non-musicians in the tolerance of clash and lapse. These results by Vuust et al. (2009) are identical to the results of the current study; thus, it seems that the disfavour of clash over lapse is innate. This is possibly due to successive prominence, which makes a rhythm difficult to understand whether or not an element is prominent: we do not tolerate the clash.

8.5 Limitations of the current study and future directions

The research reported here suggests that subjective rhythmisation can lead to perceptual illusion, but that subjective rhythmisation and the effects of the dynamic attending theory are culturally dependent. Examining this finding with experiments specifically design to address would be a valuable direction for future research. In the previous section, the possibility that English and Russian speakers could mistake test stimuli of binary rhythm for binary rhythm was discussed. However, it is unclear why this kind of illusionary binary rhythm occurred in binary rhythm trials and not in the non-binary rhythm trials, nor why it occurred only in the English and Russian speaker groups. It would be beneficial to conduct further experiments to consider whether Japanese speakers have a tendency to perceive binary rhythm as non-binary rhythm phrase more often than English or Russian speakers, and whether English or Russian speakers mistake non-binary rhythm phrase with binary rhythm phrase.

In addition, it would be necessary to consider other factors that can affect rhythm perception. One of the examples is a genetic component mentioned by Theusch and Gitschier (2011): the absolute pitch, which was thought to be culturally acquired, also has a genetic component. This suggests that genetic factors may also affect our perception of rhythm.

Musicians' accuracies were higher in non-binary rhythm in musical stimuli than those of non-musicians. Importantly, the musical experience did not enhance the accuracy of each violation type (clash, lapse, and SP) or rhythm type (binary and non-binary rhythm), except for the accuracy of non-binary rhythm in musical stimuli. This suggests that it has a smaller influence on rhythm perception than that os linguistic experience. However, it would be necessary to examine the reasons why musicians' accuracy for the non-binary rhythm of the musician was higher. In the current study, the musical stimuli were pieces of tonal music accompanied by base notes. This tonality and accompaniment were a major difference between the musical stimuli and other stimuli, in the sense that prominences in musical stimuli were underlined by additional notes. Whether the accompaniment or tonality makes a musician's ability to detect non-binary rhythm better, is something that needs further examination.

In addition, musicians' ratings for binary rhythm tended to be lower than those of non-musicians, possibly because they rated rhythmic differences in binary and non-binary rhythm based on a similar criterion (i.e. musicians rated rhythmic violation depending on whether the violation was due to a single elements or two elements), while non-musicians rated rhythmic difference depending on how they felt. The results show that the differences between ratings for binary and non-binary rhythm by musicians were smaller than those for non-musicians. It would be necessary to conduct further experiments to examine the hypothesis that musicians tend to rate rhythmic differences similarly both in binary and non-binary rhythms.

In the current study, influences of native language and musical experience are not entirely clear, as Russian musicians were not recruited; however, the fact that there was no statistical difference in rhythmic perception between English musicians and Japanese musicians provides clear evidence that speech rhythm takes precedence over musical training in one's ability to detect rhythm. Comparing Russian musicians and non-musicians would strengthen these conclusions. Similarly, it would be helpful to also test speakers of languages that are considered syllable-timed, such as French. If we can find a connection between linguistic rhythm and rhythm perception, this would support the idea that native language affects rhythm perception.

Another limitation in the current study was the broad definition of musician. Depending on the genre or eras of work they specialise in, musicians' rhythmic perceptions can differ. Similarly, a singer's perception and a percussionist's perception of rhythm can differ significantly. For instance, choir works are rhythmically uncomplicated to allow for easy synchronisation, while musical works for percussive instruments can be rhythmically complicated. It might be beneficial to compare the rhythmic perception of percussionists with that of choir singers to better understand the role of musical training in rhythm perception. Another factor that should also be considered is the age at which musicians started to learn music.

Even if the results seem to show that Japanese speakers are more sensitive to non-binary rhythm than Russian and English speakers, due to the familiarity with non-binary rhythm, it is not clear whether Japanese speakers are also sensitive to non-binary rhythms in which the durations of each element are variable. For instance, as shown in Chapter 2, Japanese speakers are more familiar than English and Russian speakers with morae, whose duration is relatively constant. This might be one reason why Japanese speakers performed well in the current study, as the sound durations in the experiments were also constant, similar to the moraic duration in Japanese, and given that Japanese moraic duration is less variable than English and Russian syllabic duration. In other words, the results may be different if the durational length of the stimuli was varied and if the stimuli consisted of a constant pattern of alternative prominence and non-prominence. Therefore, it would be helpful to hold similar experiments to the current study with durational variations introduced in the stimuli.

Although there are some limitations, this thesis:

a. Provided evidence on Japanese rhythm structure, showing that a subtle trochaic pattern based on accented syllables is used.
b. Showed that linguistic experience affects how rhythm is processed independent of modality: speakers of languages with even rhythms, like English and Russian, are less able to detect rhythmic irregularities than speakers of languages where rhythmic irregularities are frequent, such as Japanese.
c. Musical experience affects how accurately rhythm is processed, but only when rhythm becomes more complicated.

References

Abercrombie, D. (1967). *Elements of general phonetics Volume 203*. Edinburgh: Edinburgh University Press.

Abramson, A. S. & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the sixth international congress of phonetic sciences*, *196*(7), 569-573. Academia Prague.

Alderete, J. D. (2013). *Morphologically governed accent in Optimality Theory*. New York: Routledge.

Amano, S. and Kondo, T. (1999). *Nihongono goitokusei* (Lexical features in Japanese). Sanseido, Tokyo.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, *32*(2), 233–250.

Arai, T. & Greenberg, S. (1997). The temporal properties of spoken Japanese are similar to those of English. *Proceedings of the Fifth European Conference on Speech Communication and Technology*, *2*, 1011–1014.

Arvaniti, A. (2007). Greek phonetics: The state of the art. *Journal of Greek linguistics*, *8*(1), 97–208.

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.

Arvaniti, A. & Rodriquez, T. 2013. The role of rhythm class, speaking rate and F0 in language discrimination. *Laboratory Phonology*, 4, 7–38.

Badanova T. A. *Word stress in the Altai language from the comparative perspective*. Doctoral dissertation, Gorno-Altaisk University, Gorno-Altaisk.

Baharloo, S., Risch, N., Gitschier, J., & Freimer, N. B. (2000). Familial aggregation of absolute pitch. *The American Journal of Human Genetics*, *67*(3), 755–758.

Barnes, J. (2007). *Phonetics and phonology in Russian unstressed vowel reduction: A study in hyperarticulation*. Ms. Boston University.

Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Dordrecht: Foris Publications.

Beckman, M. E. & Edwards, J. (1994). Articulatory evidence for differentiating stress categories, in Phonological Structure and Phonetic Form. *Papers in Laboratory Phonology*, *3*, 1–33.

Beddor, P. S. & Strange, W. (1982). Cross-language study of the oral nasal distinction, *Journal of Acoustic Society of America*, *71*, 1551–1561.

Bekku, S. (1977). *Nihongo no Rizumu (Japanese rhythm)*. Tokyo: Kodansha.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human perception and performance*, *32*(1), 97–103.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de phonétique appliquée*, *91*(93), 99–130.

Bethin, C. (1998). *Slavic prosody: language change and phonological theory*. Cambridge: Cambridge University Press.

Bharucha, J. & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, *13*(1), 63–102.

Bion, R. A., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and speech*, *54*(1), 123–140.

Bloch, B. (1950). Studies in colloquial Japanese IV phonemics. *Language*, *26*(1), 86–125.

Boersma, P., & Weenink, D. (2011). *PRAAT: Doing phonetics by Computer (ver. 5.2. 24)*. Amsterdam: Institut for Phonetic Sciences.

Bohn, O. S. & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception, Appl. *Psycholinguistics*, *11*, 303–328.

Bolinger, D. (1958). A theory of pitch accent in English, *Word*, *14*, 109–49.

Bolinger, D. (1965). *Pitch accent and sentence rhythm, Forms of English: Accent, morpheme, order*, Cambridge, MA: Harvard University Press.

Bondarko, L. 1977. *Zvukovoj stroj sovremennogo russkogo jazyka (Sound system of modern Russian language)*. Moscow: Prosveshenie.

Bondarko, L.V. (1981). *A phonetic description of language and a phonological description of speech* Leningrad: Leningrad University Press.

Bondarko, L. V. (1998). *Fonetika sovremennogo russkogo jazyka (Phonetics of modern Russian language)*. St. Petersburg: St. Petersburg University Press.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & psychophysics*, *61*(5), 977–985.

Bybee, J. L., Chakraborti, P., Jung, D., & Scheibman, J. (1998). Prosody and segmental effect some paths of evolution for word stress. Studies in Language. *International Journal sponsored by the Foundation 'Foundations of Language'*, *22*(2), 267–314.

Campbell, N. & Sagisaka, Y. (1991). Moraic and syllable-level effects on speech timing. (Onsei taimingu ni mirareru moora to onsetsu no eikyou ni tsuite), *Acoustic Society Japan (90-107)*, 35–40.

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of speech, language, and hearing research*, *57*(4), 1468–1479.

Clayton, M. R. (1996). Free rhythm: ethnomusicology and the study of music without metre. *Bulletin of the School of Oriental and African Studies*, *59*(2), 323–332.

Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in medicine*, *16*(20), 2349–2380.

Crosswhite, K., & Jun, A. (2001). *Vowel reduction in optimality theory*. London: Psychology Press.

Crosswhite, K., Alderete, J., Beasley, T., & Markman, V. (2003). Morphological effects on default stress in novel Russian words. *Proceedings of West Coast Conference on Formal Linguistics*, *22*, 151–164.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3-4), 133–142.

Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting*, *5*(1), 1–23.

Dawe, L. A., Platt, J. R., & Racine, R. J. (1995). Rhythm perception and differences in accent weights for musicians and nonmusicians. *Perception & psychophysics*, *57*(6), 905–914.

Dellwo, V., Karnowski, P., & Szigeti, I. (2006). Rhythm and speech rate: A variation coefficient for deltaC, *Language and language processing: Proceedings of the 38th linguistic colloquium*, 231–241.

de Manrique, A. M. B. & Signorini, A. (1994). Phonological awareness, spelling and reading abilities in Spanish-speaking children. *British Journal of Educational Psychology*, *64*(3), 429–439.

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception: An Interdisciplinary Journal*, *8*(4), 335–347.

Deutsch, D., Henthorn, T., & Dolson, M (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, *21*, 339–356.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, *119*(2), 719–722.

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics*, *53*(1), 25–33.

Dupoux, E., Kakehi, H., Hirosi, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing deafness in French? *Journal of Memory and Language*, *36*(3), 406–421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). *Journal of Acoustic Society America*, *110*(3), 1606–1618.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*(2), 682–706.

Eady, S. J. & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America*, *80*(2), 402–415.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: discrimination of the /r-l/ distinction by young infants. *Perception & Psychophysics*, *18*(5), 341–347.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English/I/and/l/accurately. *Language and Speech, 38*(1), 25–55.

Fraisse, P. (1963). *The psychology of time Harper and Row.* New York: Harper & Row Press.

Fraisse, P. (1982). Rhythm and tempo; in Deutsch. *The psychology of music*, 149–180. New York: Academic Press.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech, 1*(2), 126–138.

Fujisaki, H., Hirose, K., & Sugito, M. (1986). Comparison of acoustic features of word accent in English and Japanese. *Journal of the Acoustical Society of Japan, 7*(1), 57–63.

Giegerich, H. J. (1985). Metrical phonology and phonological structure. German and English. *Cambridge Studies in Linguistics London, 43*, 1–301.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds ‘l’ and ‘r’. *Neuropsychologia, 9*: 317–323.

Gouskova, M. (2010). The phonology of boundaries and secondary stress in Russian compounds. *The Linguistic Review, 27*(4), 387–448.

Gouskova, M. & Roon, K. (2013). Gradient clash, faithfulness, and sonority sequencing effects in Russian compound stress. *Laboratory phonology, 4*(2), 383–434.

Grabe, E. & Warren, P. (1995). Stress shift: do speakers do it or do listeners hear it. *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, ed. by B. Connell and A. Arvaniti. Cambridge: Cambridge University Press, 95–110.

Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (2001). Early childhood music education and predisposition to absolute pitch: teasing apart genes and environment. *American journal of medical genetics, 98*(3), 280–282.

Gussenhoven, C. (1991). The English rhythm as an accent deletion rule. *Phonology, 8*, 1–35.

Han, M. S. (1962). The feature of duration in Japanese. *Study Sounds, 10*, 65–80.

Hannon, E. E. & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences, 102*(35), 12639–126.

Haraguchi, S. (1991). *A Theory of Stress and Accent*. Dordrecht: Foris Press.

Hasegawa, Y. & Hata, K. (1992). Fundamental frequency as an acoustic cue to accent perception. *Language and Speech, 35*(2), 87–98.

Hattori, S. (1961). Prosodeme, syllable structure and laryngeal phonemes. *Studies in Descriptive and Applied Linguistics*, 1–27.

Hattori, S. (1980). Prosodeme, syllable structure, and laryngeal phonemes. *Linguistics of Japan*, 186–218.

Hay, J. S. & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics, 69*(1), 113–122.

Hayashi, O. (1982). *Zusetsu nihon-go (Japanese dictionary)*. Tokyo: Kadokawa Press.

Hayashi, A. & Mazuka, R. (2017). Emergence of Japanese infants' prosodic preferences in infant-directed vocabulary. *Developmental psychology, 53*(1), 28–39.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry, 15*, 33–74.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Henthorn, T. & Deutsch, D. (2007). Ethnicity versus early environment: comment on ‘Early childhood music education and predisposition to absolute pitch: teasing apart genes and environment’ by Peter K. *Journal of Medical Genetics, 143*, 102–103.

Hogg, R., Hogg, R. M., Hogg, R. M., & McCully, C. B. (1987). *Metrical phonology: A course book*. Cambridge: Cambridge University Press.

Howat, R. (2009). *The Art of French Piano Music: Debussy, Ravel, Fauré, Chabrier*. Connecticut: Yale University Press.

Huron, D. & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception: An Interdisciplinary Journal, 13*(4), 489–516.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology 23*, 225–257.

Hyman, L. M. (2015). Positional Prominence vs. Word Accent: Is there a difference?. *UC Berkeley PhonLab Annual Report, 11*(11).

Hyogaki, M. (1962). *Kinki Hougen Sousetsu (Kinki Dialect)*. Tokyo: Sanseido Press.

Inoue, T. (1998). *Nihon-go Watching (Japanese Observation)*. Tokyo: Iwanami Shoten Press.

Iversen, J. R., Patel, D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124*(4), 2263–2271.

Jeon, H. S. & Arvaniti, A. (2016). Rhythmic grouping in English, Greek and Korean: testing the iambic-trochaic law. *Proceedings of Speech Prosody*, 1134–1138.

Jinbo, K. (1980). Kokugo no onseijou no tokushitsu (The top phonetic characteristics of Japanese). *Nihon no gengogaku (Linguistics of Japan), 5*, 5–15.

Jones, D. & Ward, D. (2010). *The phonetics of Russian*. Cambridge: Cambridge University Press.

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological review, 83*(5), 323–345.

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics, 32*(3), 211–218.

Jun, J., Kim, J., Lee, H., & Jun, S. A. (2006). The prosodic structure and pitch accent of Northern Kyungsang Korean. *Journal of East Asian Linguistics, 15*(4), 289–317.

Karpeles, M. (1951). Some reflections on authenticity in folk music. *Journal of the International Folk Music Council, 3*, 10–16.

Kato, H. (1999). *Perceptual characteristics of temporal structures in speech: Towards objective assessment of synthesis rules.* Doctoral dissertation, Kobe University, Hyogo.

Kato, H., Tsuzaki, M., & Sagisaka, Y. (1997). Acceptability for temporal modification of consecutive segments in isolated words. *The Journal of the Acoustical Society of America 101*(4), 2311–2322.

Kawahara, S. (2015). The phonetics of obstruent geminates, sokuon. *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology.* Berlin: Mouton de Gruyter.

Kijak, A. M. (2009). *How stressful is L2 stress? A cross-linguistic study of L2 perception and production of metrical systems*. Doctoral dissertation, Utrecht University, Utrecht.

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic inquiry, 10*(3), 421–441.

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of basic and clinical physiology and pharmacology, 12*(2), 125–144.

Kitahara, M. (2001). *Category structure and function of pitch accent in Tokyo Japanese*. Doctoral dissertation, Indiana University, Bloomington.

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J.T. (2003). Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological Science, 14*, 201–206.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America, 118*(2), 1038–1054.

Kondaurova, M. V. & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America, 124*(6), 3959–3971.

Krumhansl, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual review of psychology, 42* (1), 277–303.

Kubota, T. (1968). Gagakukyokuno Gosenkifukanikansuru Mondaito Shian (Notation of Gagaku music). *Soaijoshi University Annual Journal, 36*, 33–56.

Kubozono, H. (1995). Constraint interaction in Japanese phonology: evidence from compound accent. *Phonology at Santa Cruz, 4*, 21–38.

Kubozono, H. (1997). Lexical markedness and variation: a non-derivational account. *Proceedings of West Coast Conf. on Formal Linguistics, 15*, 273–287.

Kubozono, H. (2008). Japanese accent. In *The handbook of Japanese linguistics, ed. by Miyagawa, S. and Saito, M.*, 165–191. New York: Oxford University Press.

Kubozono, H. (2015). *Handbook of Japanese phonetics and phonology (Vol. 2).* Berlin: Walter de Gruyter GmbH & Co KG.

Kurisu, K. (1994). Further evidence for bi-moraic foot in Japanese. *Proceedings of the 1994 International Conference on Spoken Language*, 367–370.

Kusumoto, K. & Moreton, E. (1997). Native language determines the parsing of nonlinguistic rhythmic stimuli. *The Journal of the Acoustical Society of America, 102*(5), 3204–3204.

Landormy, P. & Wager, W. (1939). Maurice Ravel. *The Musical Quarterly, 25*(4), 430–441.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Child Psychology, 20*, 215–225.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lavitskaya, Y. (2015). *Prosodic structure of Russian: A psycholinguistic investigation of the metrical structure of Russian nouns*. Doctoral dissertation, University of Konstanz, Konstanz.

Lea, W. A. (1974). Prosodic aids to speech recognition: A general strategy for prosodically-guided speech understanding. *Univac Report PX10791*, 43–56.

Lerdahl, F. & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception: An Interdisciplinary Journal, 1*(2), 229–252.

Liberman, M. & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*(2), 249–336.

Lin, C. Y., Wang, M., Idsardi, W., & Xu, Y. (2014). Stress processing in Mandarin and Korean second language learners of English. Bilingualism: *Language and Cognition, 17*(2), 316–346.

Ling, J. (1997). *A history of European folk music*. University Rochester Press.

Lisker, L. & Abramson, A. S. (1967). The voicing dimension: Some experiments in comparative phonetics. *Haskins Laboratories Status Report on Speech Research, 11*, 9–15.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the acoustical society of America*, *94*(3), 1242–1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874–886.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English. *Language and Speech*, *43*, 377–401.

Lukyanchenko, A., Idsardi, W. J., & Jiang, N. (2010). Opening your ears: The role of L1 in processing of nonnative prosodic contrasts. *Proceedings of the Second Language Research Forum 2010*, 50–62.

MacKinnon, N. (1993). *The British folk scene: Musical performance and social identity*. London: Open University Press.

Maekawa, K. (2003). Corpus of Spontaneous Japanese: Its design and evaluation. *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, 7–12.

Masuda, T. & Nisbett R.E. (2001) Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81*, 922–936.

Mazuka, R., Kondo, T., & Hayashi, A. (2008). Japanese mothers' use of specialized vocabulary in infant-directed speech: infant-directed vocabulary in Japanese. *The origins of language*, *6*, 39–58.

McKone, E., Davies, A. A., Fernando, D., Aalders, R., Leung, H., Wickramariyaratne, T., & Platow, M. J. (2010). Asia has the global advantage: Race and visual attention. *Vision research, 50*(16), 1540–1549.

Morton, J. & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech, 8*(3), 159–181.

Mills, M. H. (1988). Perceived stress and the rhythmical organisation of the utterance in Colloquial Russian. *Russian Language Journal/Русский язык, 42*(141), 51–65.

Mueller, R. (1986). Javanese Influence on Debussy's Fantaisie and beyond. *Nineteenth-Century Music, 10*(2), 157–186.

Nakatani, L. H. & Aston, C. H. (1978). *Acoustic and linguistic factors in stress perception*. Ms., ATT Bell Laboratories

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance, 24*(3), 756–772.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*(1), 1–19.

Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41*(1), 233–243.

Nelson, S. G. (2017). Court and religious music (2): music of gagaku and shōmyō. *In The Ashgate research companion to Japanese music*, ed. by McQueen, A., Tokita, S, & Hughes, D, New York: Routledge Publishing, 75–102.

Nespor, M. & Vogel, I. (1989). On clashes and lapses, *Phonology, 6*, 69–116.

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. *The Blackwell companion to phonology*, Oxford: Wiley-Blackwell, 1–13.

Neustupný, J. V. (1966). Is the Japanese accent a pitch accent?. *Onsei Gakkai Kaiho (Journal of Acounstic Conference), 121*, 230–239.

Nisbett, R.E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition, *Psychological Review, 108*, 291–310.

O'Connor, J. D. (1965). *The perception of time intervals.* London: Phonetics Laboratory, University College London.

Otake, T. 1988. A cross linguistic contrast in the temporal compensation effect. *The Bulletin*, *45*, 14–19.

Otake, T. 1989. Counter evidence for mora timing. *Proceedings of the Sixteenth LACUS Forum*, 313–322.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 258–278.

Padgett, J. & Tabain, M. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica, 62*(1), 14–54.

Palmer, C. & Krumhansl, C. (1990). Mental representations of musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 728–741.

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

Peperkamp, S. & Dupoux, E. (2002). A typological study of stress ‘deafness’. *Papers in laboratory phonology, 7*, 203–240.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America, 89*(6), 2961–2977.

Polka, L. & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421–435.

Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for Mora Timing in Japanese'. *Journal of the Acoustic Society of America, 81*, 1574–1585.

Poser, W. J. (1990). Evidence for foot structure in Japanese. *Language, 66*, 78–105.

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication, 54*(6), 681–702.

Prince, A. (1983). Relating to the grid. *Linguistic Inquiry, 14*(1), 19–100.

Prokhorov, V. (2002). *Russian folk songs: Musical genres and history*. Maryland: Scarecrow Press.

Quené, H. & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62*(1), 1–13.

Ramsey, S. R. (1991). Proto-Korean and the origin of Korean accent. *Studies in the historical phonology of Asian languages, 77*, 215–232.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. *Proceedings of speech prosody 2002,* 115–120.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292.

Roach, P. (1982). On the distinction between 'stress-timed'and 'syllable-timed'languages. *Linguistic controversies*, Ed. by Crystal, D., London: Arnold, 73–79.

Rosen, E. R. (2001). *Phonological processes interacting with the lexicon: Variable and non-regular effects in Japanese phonology*. Doctoral dissertation, University of British Columbia.

Sagisaka, Y. (1984). Phoneme duration control for speech synthesis by rule. *IEICE Trans. Fundamentals (Japanese Edition), 67*(7), 629–636.

Sakano, N. (1996). *Shichi-go-cho no Nazo wo Toku (Solving the Mystery of the Seven Five Meter)*. Tokyo: Taishukan Shoten.

Sato, Y. (1993). The durations syllable-final nasals and mora hypothesis in Japanese. *Phonetica, 50*, 44–67.

Sato, Y. (1995). The Mora Timing in Japanese--A Positive Linear Correlation between the Syllable Count and Word Duration. *Onsei Gakkai Kaiho (Annual Report of Phonetic Society), 209,* 40–53.

Scherr, B. (1980). Russian and English Versification: Similarities, Differences, Analysis. *Style*, 353–378.

Selkirk, E. O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure.* Cambridge MA: MIT Press.

Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics, 22*(4), 357–388.

Shattuck-Hufnagel, S. & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research, 25*(2), 193–247.

Shen, Y. & Peterson, G. G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics - Occasional papers, 9*, 1–36.

Sluijter, A. M. & van Heuvenn, V. J. (1996, October). Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceeding of the Fourth International Conference on Spoken Language Processing. ICSLP 1996, 2*, 630–633.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Ohman, S. E. G. (1969). Crosslanguage study of vowel perception. *Language and Speech, 12*, 1–23.

Studwell, W. E. (1995). *The Christmas Carol Reader*. London: Psychology Press.

Sugito, M. (1972). Ososagari-koo: dootai-sokutei ni yoru nihongo akusento no kenkyuu (Delayed pitch fall: an acoustic study). *Shoin Joshi Daigaku Ronshuu (Annual Report of Shoin University), 10*, 201–229.

Sugiyama, Y. (2011). *The production and perception of Japanese pitch accent*. Cambridge: Cambridge Scholars Publishing.

Svetozarova, N. (1998). I*ntonation in Russian, Intonation Systems*. 261–274. Cambridge: Cambridge University Press.

Tajima, K. (1998). *Speech rhythm in English and Japanese*. Doctoral dissertation, Indiana University, Bloomington.

Tanaka, S. & Kubozono, H. (1999). *Introduction to Japanese pronunciation: Theory and practice*. Tokyo: Kuroshio.

Team, R. C. (2013). R: *A language and environment for statistical computing*.

Temperley, D. (2018). *The Musical Language of Rock*. Oxford: Oxford University Press.

Theusch, E. & Gitschier, J. (2011). Absolute pitch twin study and segregation analysis. *Twin research and human genetics, 14*(2), 173–178.

Thompson, W. F. & Cuddy, L. L. (1989). Sensitivity to key change in chorale sequences: A comparison of single voices and four-voice harmony. *Music Perception: An Interdisciplinary Journal, 7*(2), 151–168.

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29*(4), 553–584.

van Heuven, V. J. & Menert, L. (1996). Why stress position bias?. *The Journal of the Acoustical Society of America, 100*(4), 2439–2451.

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). *Predictive coding of music–brain ratings to rhythmic incongruity. cortex, 45*(1), 80–92.

Warner, N. & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica 58*(1), 1-25.

Wasow, T., Levy, R., Melnick, R., Zhu, H., & Juzek, T. (2015). Processing, prosody, and optional to. *Explicit and implicit prosody in sentence processing*. Cham: Springer. 133–158.

Weitzman, R. S. (1970). Word accent in Japanese. *Management Information Services, 9*, 45–65.

Wenk, B. J. & Wioland, F. (1982). Is French really syllable-timed?. *Journal of phonetics*, *10*(2), 193–216.

Werker, J. F. & Lalonde, C. (1988). Cross-language speech perception: initial capabilites and developmental change. *Developmental Psychology, 24*, 672–683.

Werker, J. F. & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percpetion & Psychophysics, 37*, 35–44.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganisation during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Woodrow, H. (1951). Time perception; in Stevens, *Handbook of experimental psychology*, 1224–1236. New York: Wiley.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115*(2), 356–361.

Yuan, J. (2010). Linguistic rhythm in foreign accent. *Proceedings of the Eleventh Annual Conference of the International Speech Communication Association,* 1848–1849.

Zlatoustova, L. V. (1953). *Phonetic nature of word stress in the Russian language*. Doctoral dissertation, Leningrad State University.

Appendix

Table 1: Linguistic stimuli, musical experience comparison group

Variable	Mean (accuracy)	Standard Error
English	1.77	0.0856
Japanese	2.13	0.0806
Regular	2.32	0.0635
Irregular	1.57	0.0635
Clash1	2.12	0.0762
Clash2	2.87	0.0762
Lapse1	1.66	0.0762
Lapse2	1.35	0.0762
Lapse3	1.75	0.0762
Musician	2.06	0.0850
Non-musician	1.84	0.0812

Table 2: Linguistic stimuli, musical experience comparison group

Variable	**Mean (z-scoresd rating)**	**Standard Error**
English	0.28	0.0600
Japanese	0.79	0.0594
Regular	0.64	0.0476
Irregular	0.44	0.0472
Clash1	0.52	0.0652
Clash2	1.14	0.0632
Lapse1	0.2	0.0645
Lapse2	-0.07	0.0652
Lapse3	0.23	0.0648
S.P. (Structure Preserving)	1.2	0.0623
Musician	0.5	0.0598
Non-musician	0.57	0.0597

Table 3: Musical stimuli, musical experience comparison group

Variable	**Mean (accuracy)**	**Standard Error**
English	2.26	0.0861
Japanese	2.75	0.0810
Regular	2.87	0.0627
Irregular	2.14	0.0627
Clash1	2.96	0.0726
Clash2	3.80	0.0726
Lapse1	1.68	0.0726
Lapse2	1.68	0.0726
Lapse3	2.41	0.0726
Musician	2.59	0.0854
Non-musician	2.42	0.0816

Table 5: Musical stimuli, linguistic experience comparison group

Variable	**Mean (z-scoresd rating)**	**Standard Error**
English	0.4012	0.0649
Japanese	0.7239	0.0634
Regular	0.7194	0.0489
Irregular	0.4056	0.0488
Clash1	0.9628	0.0613
Clash2	1.5629	0.0600
Lapse1	-0.0555	0.0621
Lapse2	0.0033	0.0611
Lapse3	0.5589	0.0625
S.P. (Structure Preserving)	0.3428	0.0600
Musician	0.5801	0.0647
Non-musician	0.5450	0.0636

Table 6: Pure-tone stimuli, linguistic experience comparison group

Variable	**Mean (accuracy)**	**Standard Error**
English	2.18	0.0733
Japanese	2.54	0.0690
Regular	2.75	0.0536
Irregular	1.97	0.0536
Clash1	2.74	0.0626
Clash2	3.44	0.0626
Lapse1	1.76	0.0626
Lapse2	1.85	0.0626
Lapse3	2.01	0.0626
musician	2.47	0.0728
Non-musician	2.24	0.0695

Table 7: Pure-tone stimuli, linguistic experience comparison group

Variable	**Mean (z-scoresd rating)**	**Standard Error**
English	0.43	0.0486
Japanese	0.73	0.0474
Regular	0.74	0.0374
Irregular	0.43	0.0374
Clash1	0.95	0.0491
Clash2	1.44	0.0481
Lapse1	0.12	0.0494
Lapse2	0.22	0.0493
Lapse3	0.35	0.0498
S.P. (Structure Preserving)	0.41	0.0481
musician	0.56	0.0481
Non-musician	0.61	0.0480

Table 7: Linguistic stimuli (musical experience comparison group)

Variable	**Mean (accuracy)**	**Standard Error**
English	1.53	0.1063
Japanese	2.15	0.1097
Russian	1.74	0.1097
Regular	2.23	0.0692
Irregular	1.39	0.0692
Clash1	1.87	0.0858
Clash2	2.65	0.0858
Lapse1	1.59	0.0858
Lapse2	1.24	0.0858
Lapse3	1.69	0.0858

Table 8: Linguistic stimuli (musical experience comparison group)

Variable	**Mean (z-scoresd rating)**	**Standard Error**
English	0.31	0.0836
Japanese	0.8	0.0912
Russian	0.48	0.0838
Regular	0.69	0.0556
Irregular	0.37	0.0547
Clash1	0.4	0.0742
Clash2	1.06	0.0714
Lapse1	0.23	0.0731
Lapse2	-0.11	0.0729
Lapse3	0.23	0.0739
S.P. (Structure Preserving)	1.39	0.0710

Table 9: Musical stimuli (musical experience comparison group)

Variable	**Mean (accuracy)**	**Standard Error**
English	2.15	0.1176
Japanese	2.68	0.1213
Russian	2.15	0.1213
Regular	2.87	0.0742
Irregular	1.78	0.0742
Clash1	2.74	0.0870
Clash2	3.63	0.0870
Lapse1	1.45	0.0870
Lapse2	1.52	0.0870
Lapse3	2.28	0.0870

Table 10: Musical stimuli (musical experience comparison group)

Variable	**Mean (z-scoresd rating)**	**Standard Error**
English	0.38	0.0786
Japanese	0.71	0.0837
Russian	0.63	0.0785
Regular	0.83	0.0513
Irregular	0.32	0.0514
Clash1	0.86	0.0686
Clash2	1.55	0.0664
Lapse1	-0.09	0.0691
Lapse2	-0.04	0.0678
Lapse3	0.56	0.0697
S.P. (Structure Preserving)	0.62	0.0666

Table 11: Pure-tone stimuli (musical experience comparison group)

Variable	**Mean (accuracy)**	**Standard Error**
English	2.01	0.1097
Japanese	2.47	0.1132
Russian	2.11	0.1132
Regular	2.67	0.0681
Irregular	1.72	0.0681
Clash1	2.66	0.0773
Clash2	3.35	0.0773
Lapse1	1.52	0.0773
Lapse2	1.63	0.0773
Lapse3	1.82	0.0773

Table 12: Pure-tone stimuli (musical experience comparison group)

Variable	**Mean (z-scoresd rating)**	**Standard Error**
English	0.48	0.0648
Japanese	0.71	0.0697
Russian	0.53	0.0651
Regular	0.81	0.0423
Irregular	0.34	0.0422
Clash1	0.99	0.0551
Clash2	1.47	0.0543
Lapse1	0.01	0.0551
Lapse2	0.09	0.0553
Lapse3	0.28	0.0555
S.P. (Structure Preserving)	0.61	0.0541

Table 13: List of uploaded stimuli

The stimuli are available here:

File name	Rhythm of first phrase	Rhythm of second phrase	Sound type
lingisotrial1	regular familiarisation	Control	Linguistic
lingisotrial2	regular familiarisation	Structure preserving	Linguistic
lingisotrial3	regular familiarisation	Lapse1	Linguistic
lingisotrial4	regular familiarisation	Lapse2	Linguistic
lingisotrial5	regular familiarisation	Clash1	Linguistic
lingisotrial6	regular familiarisation	Clash2	Linguistic
lingisotrial7	regular familiarisation	Lapse3	Linguistic
lingNONisotrial1	irregular familiarisation	Control	Linguistic
lingNONisotrial2	irregular familiarisation	Structure preserving	Linguistic
lingNONisotrial3	irregular familiarisation	Lapse1	Linguistic
lingNONisotrial4	irregular familiarisation	Lapse2	Linguistic
lingNONisotrial5	irregular familiarisation	Clash1	Linguistic
lingNONisotrial6	irregular familiarisation	Clash2	Linguistic
lingNONisotrial7	irregular familiarisation	Lapse3	Linguistic
musisotrial1	regular familiarisation	Control	Musical
musisotrial2	regular familiarisation	Structure preserving	Musical
musisotrial3	regular familiarisation	Lapse1	Musical
musisotrial4	regular familiarisation	Lapse2	Musical
musisotrial5	regular familiarisation	Clash1	Musical
musisotrial6	regular familiarisation	Clash2	Musical
musisotrial7	regular familiarisation	Lapse3	Musical
musNONisotrial1	irregular familiarisation	Control	Musical
musNONisotrial2	irregular familiarisation	Structure preserving	Musical
musNONisotrial3	irregular familiarisation	Lapse1	Musical
musNONisotrial4	irregular familiarisation	Lapse2	Musical
musNONisotrial5	irregular familiarisation	Clash1	Musical
musNONisotrial6	irregular familiarisation	Clash2	Musical
musNONisotrial7	irregular familiarisation	Lapse3	Musical

puretone50-150isotrial1	regular familiarisation	Control	Tonal
puretone50-150isotrial2	regular familiarisation	Structure preserving	Tonal
puretone50-150isotrial3	regular familiarisation	Lapse1	Tonal
puretone50-150isotrial4	regular familiarisation	Lapse2	Tonal
puretone50-150isotrial5	regular familiarisation	Clash1	Tonal
puretone50-150isotrial6	regular familiarisation	Clash2	Tonal
puretone50-150isotrial7	regular familiarisation	Lapse3	Tonal
puretone50-150NONisotrial1	irregular familiarisation	Control	Tonal
puretone50-150NONisotrial2	irregular familiarisation	Structure preserving	Tonal
puretone50-150NONisotrial3	irregular familiarisation	Lapse1	Tonal
puretone50-150NONisotrial4	irregular familiarisation	Lapse2	Tonal
puretone50-150NONisotrial5	irregular familiarisation	Clash1	Tonal
puretone50-150NONisotrial6	irregular familiarisation	Clash2	Tonal
puretone50-150NONisotrial7	irregular familiarisation	Lapse3	Tonal
puretone150-50isotrial1	regular familiarisation	Control	Tonal
puretone150-50isotrial2	regular familiarisation	Structure preserving	Tonal
puretone150-50isotrial3	regular familiarisation	Lapse1	Tonal
puretone150-50isotrial4	regular familiarisation	Lapse2	Tonal
puretone150-50isotrial5	regular familiarisation	Clash1	Tonal
puretone150-50isotrial6	regular familiarisation	Clash2	Tonal
puretone150-50isotrial7	regular familiarisation	Lapse3	Tonal
puretone150-50NONisotrial1	irregular familiarisation	Control	Tonal
puretone150-50NONisotrial2	irregular familiarisation	Structure preserving	Tonal
puretone150-50NONisotrial3	irregular familiarisation	Lapse1	Tonal
puretone150-50NONisotrial4	irregular familiarisation	Lapse2	Tonal
puretone150-50NONisotrial5	irregular familiarisation	Clash1	Tonal
puretone150-50NONisotrial6	irregular familiarisation	Clash2	Tonal
puretone150-50NONisotrial7	irregular familiarisation	Lapse3	Tonal

Participant Information Sheet used for English participants

Participant Information Sheet

English Language and Linguistics Dissertation – Researcher: Sumio Kobayashi

You have been invited to take part in this research study for my dissertation in English Language and Linguistics at the University of Kent. You have been asked to take part because you fall within the parameters of the study, which are as follows:

- 30 British musicians and 30 Japanese musicians
- 30 British non-musicians, 30 Japanese non-musicians, 30 Russian non-musicians

Purpose

The purpose of the study is to examine the perceptual difference of rhythm between British musicians, British non-musicians, Japanese musicians and Japanese non-musicians. Based on a study demonstrates that familiar rhythm in music affects the perception of rhythm, a hypothesis that the perceptions of Japanese and British people are affected by the linguistic and musical rhythm in their countries is examined through three experiments. Also, influence of musical experience on rhythmic perception is tested.

Expected Participation

Participants are supposed to attend four experiments that last approximately 10 minutes each.

Description of procedures

The task is to rate (from 1 to 6) the rhythmic difference of two successive sound files. '1' suggests that the rhythm of these two sound files is identical. '6' indicates that the rhythm is eminently different. The first sound file is called familiarisation stimulus and second file is test stimulus. There are six test stimuli per one familiarisation stimulus. Depending on the test stimuli, the degree of rhythmic difference between the familiarisation stimulus and test stimulus can be big or small. First experiment is designed to examine perceptual difference of music. In this experiment, familiarisation and test stimuli are musical sound file. At the second experiment, linguistic stimuli are used to consider the perceptual difference of linguistic rhythm. Lastly, in the third and the fourth experiments, pure tone stimuli, beep sound, is used. There is no information of consonant, vowel, and pitch and instrumental sonority in this pure tone stimuli, which is desirable to test purely rhythmic perception. Before the main experiment, participants are instructed through practice trials. The stimuli in the practice experiment are different from main experiment, but the degree of the difference between familiarisation stimuli and test stimuli, and procedure are identical to those of main experiment. The result is automatically saved in a computer through software.

Participation

Participation is voluntary. Participants can deny the attendance and usage of data at any time. They can contact the researcher, Sumio Kobayashi, any time as needed. Personal information is not used throughout the research.

Risks and Benefit

There is no potential risk in this study. Although there is no direct benefit to participate in the experiments, you will know your ability in rhythmic perception.

Anonymity

The personal information of each participant is not saved. After the experiment, the name of participant is replaced by abbreviation of group name, such as J.N.M. -01 (Japanese non-musician-01).

After finishing the study, data in which the personal information is removed will be stored at the linguistic lab. However, in the case in which a participant disagree with the storage, the participant can state that it is desirable to delete the data after the research and can contact the researcher, Sumio Kobayashi, to ask the deletion of data after the experiments.

Contact Details:

Faculty of Humanities,

University of Kent,

Canterbury,

Kent, CT2 7NR

Humanities Department Email: hsugo@kent.ac.uk

The study is voluntary. You can ask any questions and withdraw at any time. Please contact me if you would like to ask or withdraw. If you would like a copy of final study, please indicate on the consent form.

Participant Information Sheet used for Japanese participants

参加者の為のインフォメーションシート

英語学言語学学位論文 – 研究者:小林純生

ケント大学の英語学言語学学位論文の研究の一環としてご案内しています。ご参加をお願いしたのは、本研究のために必要な参加者の条件に合致しているためです。その条件は以下のものです：

- 30人のイギリス人音楽家、 30人の日本人音楽家
- 30人のイギリス人非音楽家, 30人の日本人非音楽家, 30人のロシア人非音楽家

目的

本研究の目的はイギリス人音楽家、日本人音楽家、イギリス人非音楽家、日本人非音楽家、ロシア人非音楽家の間にリズムの知覚に対する差異があるかを調査することです。慣れ親しんでいるリズムに知覚が鋭敏になるという先行研究を踏まえ、イギリス人、日本人、そしてロシア人はその母国語と音楽の経験によってリズムが異なると仮説をたてることができ、その仮説を実験を通して証明します。

参加について

参加者の皆様には、四つの実験、ひとつあたりおよそ10分程度のもの、に参加していただきます。

実験について

本研究でお願いしたいのは二つの音声ファイルを聞いていただき、それらの間にリズムの違いがどれくらいあったかを答えていただくことです。１という回答は二つのファイルの間にリズムの違いが無かったという回答になります。また、６という回答は、それら二つのリズムは極めて異なっていたという回答になります。最初に再生されるファイルはファミリアリゼーション刺激と呼ばれます、また二つ目のファイルはテスト刺激と呼ばれます。テスト刺激には六種類のものがあり、それぞれのファミリアリゼーション刺激毎に異なる六種類のテスト刺激があります。テスト刺激によって、二つのリズム間の違いが大きいものや小さいものがあります。まずは音楽の知覚を検討するために、音楽の音を使った実験があります。そしてもう一つの実験では言語のリズム知覚を調査するために、言語音が用いられます。そしてその他に純音刺激をつかって実験が用意されており、ビープ音が用いられます。この純音刺激では子音、母音、音高や響きといったものが無く、リズムのみを検討するために作られています。本実験を開始する前に、練習用の実験が用意されていて、そちらにまずは参加し、概要を理解していただけたらと思います。この練習用の実験では、リズムが本実験とは異なりますが、音声ファイル間のリズムの差異の大きさは本実験と同様ですし、操作方法や実行方法も本実験と同様です。回答内容に関しては、自動的に保存されます。

倫理的配慮について

本実験の参加は任意であり、強制ではありません。実験参加や実験への出席、データの使用をいつでも拒否できます。またいかなる場合においても、参加者は小林純生に連絡でき、上記の事項を申請できます。個人情報は如何なる場合も用いることはありません。

参加におけるリスクについて

There is no potential risk in this study. Although there is no direct benefit to participate in the experiments, you will know your ability in rhythmic perception.

潜在的なものも含め、参加によるリスクはありません。直接的な恩恵も参加に関して無いと言えるかも知れませんが、研究結果をお伝えすることができます。

匿名性

参加者の個人情報は一切保存されません。実験後参加者の氏名はJ.N.M. -01 (Japanese non-musician-01)といったコードに置き換えられます。匿名性が保たれたデータを研究室に保存します。しかしながら、参加者の方がそのデータの保存に賛成しかねるという場合、小林純生まで連絡し、データの削除を求めて下さい。

連絡先について：

Faculty of Humanities,

University of Kent,

Canterbury,

Kent, CT2 7NR

Humanities Department Email: hsugo@kent.ac.uk

この研究への参加はあくまでも任意です。いかなる質問や要求、データの削除にも対応します。いつでもご連絡ください。また研究結果に関しましてもお伝えできますので、そういった際もご連絡してください。

Participant Information Sheet used for Russian participants

Информационный лист участника

Диссертация по английскому языку и лингвистике – Исследователь: Sumio Kobayashi

Вы были приглашены принять участие в этом исследовании для моей диссертации по английскому языку и лингвистике в Университете Кента. Вас попросили принять участие, потому что вы попадаете в параметры исследования, а именно:

- 20 британских музыкантов и 20 японских музыкантов
- 20 британских не музыкантов, 20 японских не музыкантов и 30 русских не музыкантов

Цель

Цель исследования - изучить разницу в восприятии ритма между британскими музыкантами, британскими не музыкантами, японскими музыкантами и японскими не музыкантами, русскими не музыкантами. Основанновываясь на исследованиях, что знакомый ритм в музыке влияет на восприятие ритма, выдвинута гипотиза о том, что восприятие японцев, британцев и русских зависит от языкового и музыкального ритма в их странах. Исследование будет проводиться с помощью трех экспериментов. Также будет исследоваться влияние музыкального опыта на ритмическое восприятие.

Требования к участникам:

Участники должны присутствовать на четырех экспериментах, которые будут длиться примерно по 40 минут.

Описание эксперемента

Задача состоит в том, чтобы оценить (от 1 до 8) ритмическую разницу двух последовательных звуковых файлов. «1» предполагает, что ритм этих двух звуковых файлов идентичен.«8» означает, что ритм совершенно другой. Первый звуковой файл(стимул) для ознакомления, а второй – тестовый. На один ознакомительный стимул приходится семь тестовых стимулов. В зависимости от тестовых стимулов степень ритмического различия между стимулом ознакомления и тестовым стимулом может быть большой или маленькой. Первый эксперимент предназначен для изучения различий восприятия музыки. В этом эксперименте ознакомительные и тестовые стимулы представляют собой музыкальный звуковой файл. Во втором эксперименте лингвистические стимулы используются, чтобы учесть разницу в восприятии языкового ритма. Наконец, в третьем и четвертом эксперименте используются чистые звуковые стимулы, звуковые сигналы. В этих звуковых стимулах нет информации о согласной, гласной, высшей и инструментальной звучности, которая желательна для проверки чисто ритмического восприятия. Перед основным экспериментом участники проходят инструктаж, и делают пробный практический эксперимент. Стимулы в практическом эксперименте отличаются от стимулов основного эксперимента, но степень различий между ознакомительными стимулами пробного эксперемента и стимулами основного эксперемента идентична. Процедура пробного эксперемента идентична процедуре основного эксперимента. Результат автоматически сохраняется на компьютере через программное обеспечение.

Участие

Участие добровольное. Участники могут отказать от участия в эксперементе в любое время. Они могут связаться с исследователем Sumio Kobayashi, когда им это нужно. Личная информация не используется для исследования.

Риски и польза

В этом исследование нет потенциального риска. Хотя нет непосредственной пользы для участника эксперемента, но вы можете узнать о своем ритмическом восприятии.

Анонимность

Персональная информация каждого участника не сохраняется. После эксперимента имя участника заменяется сокращением имени группы, например: J.N.M. -01 (Japanese non-musician-01).

После окончания исследования данные, в которых будет удалена личная информация, будут храниться в лингвистической лаборатории. Однако в случае, если участник не согласен с храннием результатов его эксперемента, он может заявить, что желает удалить данные после исследования, для этого он должен связаться с исследователем Sumio Kobayashi, чтобы попросить удалить данные после экспериментов.

Контактная информация:

Faculty of Humanities,

University of Kent,

Canterbury,

Kent, CT2 7NR

Humanities Department Email: hsugo@kent.ac.uk

Участие добровольное. Вы можете задать любые вопросы в любое время. Пожалуйста, свяжитесь со мной, если у вас возникли вопросы или вы хотите отозвать результаты вашего эксперемента. Если вам нужна копия итоговогов исследования, укажите это в форме согласия.