# The Relationships between Sound Sensitivity, English Prosody Processing, and English Listening Comprehension

TABATA Megumi

A Thesis

Submitted

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Graduate School of Languages and Cultures,

Nagoya University

March 2016

# Abstract

From the late 20th century onwards, the perception and production of English prosody have been recognized as crucial factors in native-like comprehension and production; however, do they affect the practical use of English? If prosody processing skill affects listening comprehension or production in English, how should EFL (English as a Foreign Language) learners acquire this skill? Fundamentally, does aptitude for sound sensitivity influence the acquisition of English prosody processing skill? Generally, sound sensitivity, comprising factors such as rhythm and tone sensitivity, is thought to be a talent that is acquired early through certain auditory training (Gordon, 1982). If auditory training does implicitly affect English comprehension and production by improving English prosody processing skill, future English learners can prepare for school English classes by honing their sound sensitivity at an early age. In this thesis, the research issues outlined above will be addressed, based on the findings of previous research.

No comprehensive empirical research has yet been conducted on the relationship between sound sensitivity, English prosody processing, and English listening comprehension and production. However, some relevant research has been conducted that partially addresses the relationship between these three variables. These research studies are as follows:

The relevant research on the relationship between sound sensitivity and general English ability includes studies on (1) the relationship between general sound processing skill and English phonological awareness, short-term memory, reading proficiency, and pronunciation (Anvari, Trainor, Woodside, & Levy, 2002; Atterbury, 1985; Call, 1985; Douglas & Willats, 1994; Lamb & Gregory, 1993; Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008; Milovanov et al., 2009; Peynircioğlu, 2002; Slevc & Miyake, 2006; Weinert, 1992). These studies are relevant to investigating the relationship between sound sensitivity and English proficiency, which will be the focus of this study.

Relevant research has also been conducted on (2) the relationship between sound processing and language prosody processing (François & Schön, 2014; Koelsch et al., 2002; Levintin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Magne, Schön, & Besson, 2006; Marques, Moreno, Castro, & Besson, 2007; Nakano, 2012; Patel & Daniele, 2003; Pesetsky, 2008; Schön, Magne, & Besson, 2004; Thompson, Schellenburg, & Husain, 2004). Finally, relevant research has been conducted on (3) the relationship between English prosody processing skill and English proficiency (Anderson-Hsieh, Johnson, & Koehler, 1992; Aoyama & Guion, 2007; Gottfried, 2007; Meerman, Kiyama, & Tamaoka, 2014; Thompson, Schellenburg, & Husain, 2003). The previous research related to these relationships will be introduced in

Chapter 2. To summarize, these research studies focus on (1) the relationship between sound sensitivity and English proficiency such as verbal memory, phonological awareness, pronunciation, reading skill, and listening comprehension; (2) the relationship between sound sensitivity and language prosody processing skills such as focusing, chunking, and intonation processing; and (3) the relationship between English prosody processing skill and English proficiency. After examining the previous research, it appeared that the broader triangle or linear relationship between general sound processing, English prosody processing, and English listening comprehension has been underexplored. In this doctoral research, this relationship will be studied in detail with the cooperation of 8–10 and 15–17 year old EFL learners. The research hypothesis on the relationships between the three variables, i.e., general sound processing, English prosody processing, and English listening comprehension, will be comparatively tested in both age groups.

In order to investigate the effectiveness of early instrumental music training on sound sensitivity, the sound sensitivity of musically trained and untrained children was compared. In this study, musically trained children were categorized as those who had been trained in Western instrumental music, including training in instruments such as the piano, flute, electronic piano, or violin, for at least a 30-minute lesson per day for more than three years before the age of nine years. Those who did not fulfill these criteria were categorized as musically untrained children.

The first objective of this dissertation is to clarify whether sound sensitivity, particularly tonal/rhythm perception and rhythm production, are related to Japanese school children's receptive/productive prosody processing skills in English, focusing on Japanese students in year 3 of elementary school, and students in years 1 and 2 of high school. The second objective is to clarify whether learners' English prosody processing skill correlates with their English listening comprehension skill. The final objective is to determine whether musically trained children have a higher sound sensitivity than untrained children, which would indicate that early music training could contribute implicitly to enhancing English listening comprehension.

In order to realize the first and second objectives, Study 1 and Study 2 were conducted with the cooperation of 128 participants in total. Seventy-five Japanese public high school students participated in Study 1. The Musical aptitude test, English prosody perception and English Prosody discrimination tests, and the Eiken[1] Grade 3 listening test were administered to the high school participants during class. Regarding the results, a Pearson's correlation analysis was conducted between all the variables used in this study. Multiple regression analysis and structural equation modeling (SEM) were also conducted.

In Study 2, fifty-three children attending an elementary school were individually administered the same Musical aptitude test, English Prosody perception test, and English Prosody discrimination test that were administered to the high school participants. In addition to these tests, the Rhythm production test and English prosody production test were administered individually, and trial versions of the Junior English Test (JET) were administered in class. With regard to Study 2, Pearson’s correlation analysis, multiple regression analysis, and SEM analysis were conducted on the data.

In Study 3, the data from the common tests for the high school participants and elementary school participants were comparatively analyzed with *t*-tests. This analysis was conducted to see if there is any difference in sound sensitivity and English prosody processing skill between 8–10 year old elementary school students and 15–17 year old high school students. By analyzing these differences, the future prospects of music and English education in elementary schools and high schools can be gauged.

In Study 4, sound sensitivity was compared between 32 musically trained children and 42 untrained children, aged from 6-12 years. All the musically trained children had been practicing musical instruments for more than 30 minutes per day for more than three years. They comprised 1 guitar student, 1 flute student, 3 electric piano students, 13 piano students, and 14 violin students.

A *t*-test was conducted to compare the musically trained and untrained participants. It was observed that musically trained children were statistically better in tone/rhythm perception. Although further evidence is required, these results indicate that early music training could help improve sound sensitivity.

In Chapter 7, the conclusions and implications for the EFL environment, limitations of the study, and suggestions for future research will be discussed. The suggestions for future research will be proposed with regard to overall English proficiency, methodology for enhancement of sound sensitivity and English prosody processing skill, and the best age for starting music and EFL education.

*Keywords:* early music training, English prosody, Japanese learners of English, listening comprehension, sound sensitivity

## Acknowledgements

I would like to express my deep appreciation to my supervisor Remi Murao for giving me insightful advice and encouragement over the past three years. I would also like to thank Shuji Ozeki for supervising my efforts to improve my English, and Katsuo Tamaoka for encouraging me and instructing me greatly in SEM analysis.

I am also immensely grateful to the following teachers: Markus Rude from Mei-writing for tutoring me on my writing and thesis statements. Masatoshi Sugiura for giving me the opportunity to present my research, reviewing the research process, and commenting critically and professionally in his seminars. Toru Kinoshita for offering me the opportunity to deliver a presentation on my research and for generously providing valuable advice, particularly on statistics. Masumi Kojima for providing me with specialist advice on the statistics used in this research. Tim Murphy for taking the time to read my thesis and for offering his insightful comments, which have enhanced my perspective. Tanomu Kashima for commenting on several details in my research and for recommending useful references for my study. Mahito Fukuda for providing moral support.

I would also like to express my gratitude to the high school and elementary school teachers and students for cooperating with me in this research. I am grateful to my friends

Yusaku Kawaguchi, Aki Goto, and Daichi Sugai for guiding me in completing my thesis. I am also grateful to the violin students and their mothers who cooperated with me in this research. I am beholden to Matthew Goldman and the international students from overseas who helped me with recordings for the test production and evaluation. I would like to express my deep appreciation to Takako Kugimiya for giving me advice on my original Rhythm production test and for devoting several hours to the test evaluation.

Finally, I would like to thank my family, parents, and the office staff for their practical support throughout the process of the doctoral degree.

For further inquiries concerning this dissertation, please e-mail me at tabatapipi@gmail.com

Megumi Tabata

Nagoya, Japan

March 2016

**Table of Contents**

**List of Tables**

**List of figures**

The Relationships between Sound Sensitivity, English Prosody Processing, and English Listening Comprehension

# Chapter 1 Introduction

## 1.1 Background of the Research

As described by Japan’s Board of Education, English is a means of communication; however, many novice learners of English may perceive spoken English words as random sounds or as a melody, and they hear these words as if they were the chatter or singing of birds. This may occur because music and language have similar roots. This is evident from the sounds made by babies crying for attachment or love; children are exposed to their mother tongue while in utero, and they spontaneously acquire this language in time. However, while in utero, they are only conscious of prosody, and not words or vocabulary. This study will focus on sensitivity to the prosodic elements of language, because children are thought to develop their language from prosodic elements toward means of communication for daily use. The purpose of learning English is not only to gain knowledge, but also to become familiar with its characteristic prosody and enjoy using it for communication.

From the author's involvement in music education as a violin student and in English education as a teacher, she has observed that advanced English learners tend to have better sound sensitivity. In particular, from the author's impression over 13 years of high school teaching experience, musically talented learners are likely to make more progress in EFL than those who do not have a special musical talent. Relevant research has been conducted on the relationship between language prosody processing skills and music processing skills (see Chapter 2) in various fields of study, such as neurological science, education, music, and linguistics. Some previous studies and measures have focused on musical and linguistic aptitude. Linguistic aptitude for foreign language acquisition is tested with, for example, the Pimsleur Language Aptitude Battery (PLAB; Pimsleur, 1966) and High-level Language Aptitude Battery (Hi-LAB; Doughty et al., 2010). PLAB is structured so as to evaluate the four factors that are significantly related to successful foreign language learning: grade point average, motivation, verbal ability, and audio ability. In Hi-LAB, audio ability accounts for nearly half of the score for the total battery of tests.

Musically talented learners seem likely to advance in English, which means that they often have a high aptitude for acquiring English proficiency. Thus, the high audio ability of musically talented learners might enable them to acquire high English competence. The audio abilities examined in these language aptitude tests are phonological awareness and phonological discrimination. Furthermore, the English prosody processing skill test, Profiling Elements of Prosodic Systems-Children (PEPS-C), which can measure the cognitive processing abilities required for English communicative skills, has been explored by Peppé and McCann (2003). The three linguistic aptitude tests described above, PLAB, Hi-LAB, and PEPS-C all measure audio ability, such as English phonological and prosodic ability.

The acquisition of adequate English prosody perception and production skills has been recognized as essential for native-like comprehension and pronunciation (Aoyama & Guion, 2007). Musical elements in language, such as rhythm and pitch, are conceptualized in combination as language prosody and have been extensively studied in relation to language acquisition (Weinert, 1992). Although the relationship between sound sensitivity and language prosodic processing has been discussed (Friedrich, 2004; Nakano, 2012), neither the procedure nor methodology for cultivating sound sensitivity as a prerequisite fundamental ability to enhance English prosody processing skill have yet been fully examined. Thus, in light of previous studies, this doctoral research proposes the hypothesis that sound sensitivity is a fundamental ability for EFL learners to acquire English prosody processing skill, and consequently, English listening comprehension skill.

## 1.2 Organization of the Thesis

In order to clarify the acquisition process from sound sensitivity via English prosody processing skill to English listening comprehension skill, the relationship between general sound processing skills such as tone and rhythm processing, the prosody perception/production of English, and English listening comprehension is targeted as the research topic. Another research topic is the effect of early music training on sound sensitivity, with reference to the previous literature on early effects on sound sensitivity as a potential language aptitude.

With regard to terms that are used frequently throughout this thesis, the following definitions apply. *Aptitude* is defined as a measure of one's potential to learn (Gordon, 1982, p. 3). Doughty et al. (2010) hypothesized that high-level language aptitude consists of inherent cognitive and perceptual abilities that can compensate for the typical post-critical period degradation in language learning capacity (p. 10). They also listed the major constructs hypothesized to underlie high-level language aptitude as memory, perceptual acuity, speed, primability, induction, pragmatic sensitivity, and fluency (p.12). *Audio ability* in language aptitude tests mostly refers to phonological sensitivity, which has been tested in phonemic discrimination or phonemic categorization (Doughty et al., 2010) and in phonetic distinction and sound symbol-association (Pimsleur, 1966). *Sound sensitivity* is defined as the acuity of tone and rhythm processing of metrically structured musical notes. It is measured as musical tone and rhythm sensitivity. The concepts of sound sensitivity and English prosody processing skill will be explained at the beginning of Chapter 2.

Further, the relevant previous research will be reviewed in the following order: (1) the relationship between music sound processing and language prosody processing, (2) the relationship between English prosody perception and production skill and English proficiency, and (3) the supposition of early prosody acquisition in EFL programs and the contribution of early music education to language acquisition. At the end of Chapter 2, the proposals and general purposes of experimental Studies 1–4 and the research questions of this thesis are described.

In Chapter 3, a description and discussion of Study 1 will be provided. Study 1 was conducted to clarify the relationship between sound sensitivity, prosody perception, and EFL listening comprehension in high school students. In Chapter 4, Study 2 will be described and discussed. Study 2 was conducted to clarify the correlation between sound sensitivity, prosody perception/production processing of English, and the English listening comprehension of elementary school children. In Chapter 5, Study 3 will be described and discussed. Study 3 was a comparative study between the data from Studies 1 and 2, in order to compare sound sensitivity and English prosody perception between elementary school students and high school students.

The underlying objective of Study 3 was to investigate how music and English education should be structured and provided from elementary school to high school, which will be discussed in Chapter 7.

In Chapter 6, Study 4 will be described. Study 4 was conducted to compare sound sensitivity between musically trained children and musically untrained children aged from 6-12, in order to observe the effect of early music lessons on sound sensitivity. In Chapter 7, conclusions and implications from Studies 1–4 will be provided.

# Chapter 2 Literature Review

## 2.1 Previous Research on Sound Sensitivity for Perception and Production of Music and Language

As mentioned in the Introduction, the relationships between sound sensitivity, English prosody processing skills, and English listening comprehension skill is one of the themes of this thesis. Before going into the details of the relationships between these variables, it is necessary to clarify whether an identical or similar process occurs when learners use prosodic cues to comprehend music and language. The author also queries the general idea that early music training is more efficient than late-started training in acquiring sound sensitivity.

In order to ascertain how sound sensitivity for language and music is acquired from birth, and the reason why early music training can be more effective, particularly with regard to acquiring absolute pitch,[3] than late-started training (Crozier, 1997), the concepts of music and language will be reviewed in the following section.

**2.1.1. Human infants' sound discrimination of language and music.** Firstly, in order to study the correlation between music and language, it is necessary to clarify what language and music are and where they come from, i.e., their concepts and roots. Language and music have been commonly described as early-acquired communication tools in the previous literature

(Gordon, 1982; Krashen, Long, & Scarcella, 1979; Trainor, 2005). This section proposes the common characteristics and history of language and music.

The first shared aspect of music and language is that human exposure to both begins before birth. According to medical research (Shimizu, 2002), even before human language acquisition, a mother’s words may reach a baby in utero, in the form of a melody. François and Schön (2014) have explained that infants have the ability to discriminate between speech sounds of several different languages and are sensitive to statistical regularities[4] in their mother language. Unfortunately, little academic research has been conducted on either prenatal mother talk or language learning. However, the effects of music in pregnancy have been presented by Shimizu (2002), who recorded how the heartbeats and movement of unborn babies changed with exposure to familiar music that had previously been sung by their mothers or played for them. This suggests that unborn babies have a sensor and react to their mothers’ songs.

Since babies are exposed to their mother tongue while in utero, they gradually acquire this language after birth. While in utero, babies are not conscious of words but of sounds, i.e., the prosody of words. Unborn babies may possess a sensitivity to the tone and rhythm of their mothers’ talk, i.e., prosody, before understanding its syntax or semantics (Ramus, 2002a). Previous studies have shown that newborn infants can categorize varieties of speech rhythm; for example, Ramus (2002a) proved that newborn infants can discriminate between the rhythms of Dutch and Japanese, using a speech re-synthesis technique to progressively decrease non-rhythmical properties of the sentences, although the effect was weaker than when intonation was present.

Each process of human language development demonstrates that language and music share a common origin as a means of communication, and that they have been developed and functionalized variously in human history, as seen in the existing thousands of languages and dozens of music scales all over the world. In human history, various kinds of music and language have evolved. In the next section, the concepts of linguistic and musical prosody will be further described, with reference to previous research.

**2.1.2. Functions of language and music.** In the era of globalization, English has been described as a means of communication by the Japanese Ministry of Education, Culture, Sports, Science and Technology (2015a; henceforth, MEXT), in its guidelines for elementary to high school education. Interpersonally, language is a means of communication through which knowledge, beliefs, and behavior are experienced. While vocalization can be produced by both humans and animals, in barking, mooing, growling, etc., verbalization is only produced by humans. Dunbar (1997) described how language evolved as an alternative to physical grooming.

In other words, physical grooming was replaced by “vocal grooming,” even if its purpose has remained the formation and maintenance of friendships or alliances.

According to Trainor (2005), music is also a communication tool, whose characteristics are rhythmic structures and pitch structures. With regard to the evolution of music, Charles Darwin proposed that it might have arisen due to sexual selection in mating calls (Darwin, 1872), while Huron (2003) suggested that music making might have arisen as a courtship behavior, to demonstrate the ability to sing well and a state of good health. Huron (2003) also supported the theory of music and social relationships, which claims that various mental disorders are linked simultaneously to social/interpersonal and musical functions. For example, in addition to high verbal abilities, individuals with Williams syndrome also exhibit high sociability and high musicality. Williams syndrome is a rare neurodevelopmental disorder characterized by a distinctive elfin facial appearance along with a low nasal bridge, an unusual ease with strangers, and developmental delay coupled with strong language skills. The syndrome was first identified in 1961 by New Zealander J. Williams. Contrarily, in spite of their high verbal abilities, people with Asperger’s-type autism can have symptoms including low sociability and low musicality.

Another theory that can be applied to both language and music acquisition is *early* *acquisition theory* or the *critical period hypothesis*. During the language critical period, humans acquire language implicitly (Long, 1990a), while in the music critical period, which continues up to age nine, they acquire musical aptitude (Gordon, 1982). The critical period hypothesis for language acquisition was theoretically supported by Krashen et al. (1979) and Slavoff and Johnson (1995), whereas the critical periods for the development of musical aptitude have been disputed. One aspect of musical aptitude, i.e., absolute pitch, is considered to be acquirable up to age six (Crozier, 1997; Takeuchi & Hulse, 1993). Early acquisition theory seems partly applicable to the acquisition of musical talent and language competence (see the citations on the critical period hypothesis later in Chapter 2).

From these theories about language and music as early-acquired communication tools, it will be worth investigating in this thesis how they function inter-connectively and their mutual effect. In the next section, the implicitly learned sound processing of language and music, or so-called *prosody*, is discussed.

**2.1.3. Discrepancy between musical and linguistic rhythm structures.** Musical rhythm follows the time-span reduction rule[5] (Jackendoff & Lerdahl, 1983, p. 146), which differs from the *prosodic rules of prominence* (Jackendoff & Lerdahl, 1983, p. 322). Time-span reduction and prosodic structure are represented by different trees comprised of either musical strings or phonological strings (p. 326). Each string is segmented into a layered hierarchy, in which each layer provides further layered segmentations. In a time-span tree of music, the opposition in each segmentation is head versus elaboration, while in a prosodic tree, the opposition is strong versus weak. In time-span reduction rules, segmentations are created by group boundaries, whereas in prosodic prominence rules, segmentations are created by word boundaries.

According to the grammar of time-span reduction, there are three rule components: Firstly, the segmentation rules (Jackendoff & Lerdahl, 1983, p. 146) produce an exhaustive hierarchical segmentation of musical surface into time-spans. Time-spans form a layered hierarchy. At the most local levels are time-spans called subgroups, those determined by metrical structure alone and by metrical structure interacting with groups. At the next levels, there are one or more levels of uncadenced group; and at the largest levels, the time-spans are all cadenced groups. These layers are fixed in order from smallest level to largest, and they cannot mix. Secondly, the well-formedness rules (Jackendoff & Lerdahl, 1983, p. 148) explain the second component of the grammar of time-reduction. They determine a set of time-span reduction trees, where every time-span has a head selected from the heads of the time-spans preceding the time-span concerned. Thirdly, the choice of head for each time-span is determined by the preference rules (Jackendoff & Lerdahl, 1983, p. 148), which are sensitive to the metrical, harmonic, voice-leading, and cadential properties of the events in the time-span.

According to Jackendoff and Lerdahl (1983), who explained the difference in rhythm and commonality between music and language, “musical events are organized around a fixed and regular metrical structure that must be maintained throughout. In language, by contrast, the rhythm is flexible and is not required to conform to any particular pattern.” They concluded:

> (T)he relation of relative prominence to metrical structure is the same in music and language, and the difference between the musical preference rules and the prosodic Rhythm Rule is a function largely of the different metrical practices of the two media. (p. 326)

Comparatively to music metrical structures, the four-level prosodic rules of prominence were explained in Jackendoff and Lerdahl (1983, p. 322): Each of the layers in the prosodic hierarchy has a characteristic rule or set of rules for determining the relationships of prominence. Starting from the innermost layer, which is the aggregation of syllables into feet, the rule for English is as follows:

1. Prominence Rule 1

   (a) In a foot that immediately contains two syllables, the first syllable is strong.

   (b) In a foot that immediately contains a foot and a syllable, the foot is strong.

2. Prominence Rule 2 (Lexical Category Prominence Role: LCPR) (Liberman & Prince, 1977, p. 270)

   In a segment immediately containing two feet, the second is strong if and only if it branches.

3. Prominence Rule 3 (Compound Stress)

   In a word immediately containing two words, the second is strong if and only if it branches into two words.

4. Prominence Rule 4 (Nuclear Stress)

   In a phonological phrase that immediately contains two phonological phrases or words, the second is strong.

As described above in each rule, a deep discrepancy exists between the rules of time-span reduction and prosodic prominence rules. In spite of the equivalence of prosodic trees to the plainest form of time-span trees, where only ordinary reduction has taken place, there are three other relationships that can be obtained between the head of a time-span and the events within it: Fusion, transformation, and cadential retention (Jackendoff & Lerdahl, 1983, p. 326). Prosodic theory has no equivalent to these three relationships.

The second difference between the two theories is plurality of branching. In time-span theory, most of the branching is binary; each time-span above the smallest level contains two time-spans of the next smallest level. Comparatively, prosodic theory has been formulated in terms of strictly binary branching; where a foot contains more than two syllables, the segmentation and resulting tree are constructed with recursive binary branching, rather than multiple branching (Liberman & Prince, 1977, p. 262).

The third difference between time-span and prosodic trees concerns the branch: A branch at any level of a time-span tree of a music piece stands for a single pitch-event, which is elaborated by all the events whose branches are subsidiary to it. In a prosodic tree, a branch is taken to stand for an entire prosodic constituent containing strong and weak parts (Liberman & Prince, 1977, p. 263). Concerning the prosodic structure of language, Liberman (1975) and Liberman and Prince (1977) have developed the concepts of *prosodic tree structures*, i.e., metrical structures. The prosodic tree developed by them expresses a syllabic structure where a head (word) separates into its end feet. Each word is separated from the head into the first feet with one foot of stressed syllables (strong) and the other foot of unstressed syllables (weak). The first foot separates several times into its final feet constituted of the minimum number of syllabic units. Liberman and Prince (1977) used the notion of syllable and foot, and Selkirk (1978) argued for the existence of prosodic categories from syllables at the smallest level to feet, word, phonological phrase, intonational phrase, and utterance at the largest level.

As described in this section, time-span theory explains the rhythm structure of music and prosodic theory explains the rhythm structure of language. The structural difference established from comparison between the two theories is summarized: At the smallest level of structure, the relationship of relative prominence to the structure is the same in both media, but at the largest level of metrical practice, the whole structures are largely different between the two.

**2.1.4 Common points in linguistic and musical prominence.** Ladd (2008) described phonological stress as distinguished in three ways: Abstract prominence relationships ("metrical strength"), concrete acoustic prominence on a particular syllable ("dynamic stress"), and the location of prominence-related intonational events ("pitch accent") (p. 58). In cases of metrical strength and pitch accent, a syllable can be metrically strong or prominent without necessarily being stressed. Comparable dissociation can be found in music: For example, dissociation occurs spontaneously between rhythmic prominence (defined by underlying beat) and melodic or dynamic prominence (defined by harmonic changes, note duration, added loudness, etc.) (p. 62). Mostly in language, syllables that are prominent in the abstract metrical structure are also phonetically stressed; however, dissociation sometimes occurs. In English and Dutch, the metrically prominent syllables with which pitch accents are associated are also dynamically stressed. However, this relationship is restrictively applied to languages such as English and Dutch. In other languages, such as French or Indonesian, pitch accents may associate with syllables that are not necessarily stressed and that may not seem prominent either.

In this thesis, under the theory that the relationship of relative prominence to the structure is the same in music and language, the assumption will be made that Western instrumental music training in youth can contribute to English prosody processing and English listening comprehension. In addition to the musical and linguistic rhythm structure and prominence discussed so far, the concept of English prosody will be discussed in the following section.

**2.1.5 Nuclear-stress placement as a key factor in rhythmic English prosody.** Among the various concepts of prosody proposed by researchers, Markus's (2005) supra-segmental phonology will first be introduced followed by Ichikawa's (2011) communicative aspect. While prosody is defined linguistically as "supra-segmental" by Markus (2005), Ichikawa (2011) focuses on the functions of prosody and defines it as "message preliminary announcement information."

As Algeo (1992) described, "Prosody goes beyond the study of phonemes and deals with such features as length, rhythm, stress, pitch, intonation, and loudness in speech" (p. 818). Inoue (2006) also describes prosody as a supra-segmental phenomenon. According to Inoue (2006), prosody consists of perceptual phenomena such as pitch, loudness, and length of linguistic sound, which correspond to aspects of physical reality such as frequency, amplitude, and duration of sound waves.

According to Markus (2005), the term prosody covers the linguistic features of musical elements of spoken language and includes three physical parameters of sounds: (a) their extension in time (length/duration); (b) their loudness/sound intensity, as caused by wave amplitudes; and (c) their pitch or pitch contour, as caused by sound frequencies. Rhythm is produced by length, loudness, and pitch when applied regularly. Stress is a term based on loudness and pitch in combination with length. Intonation is largely created by modifying pitch and loudness in their temporal extension.

Markus (2005) described the prosodic nucleus as the main and compulsory unit in “a tone group or speech group” marked by its suggestive internal pitch contour. He describes how both English and German tone units in a sentence have a compulsory nucleus, which comes after the optional pre-head, head, and is followed by a body and tail. He describes the compulsory nature of the nucleus, compared to all the other optional elements in a tone group. In general English statements, the head is considerably higher in pitch than the nucleus, and the pitch falls within the body in steps from one stressed syllable to the next. The head and nucleus are thus identified by this “regular pattern of the flight of stairs” (p. 116).

Algeo (1992) considers prosodic nuclei in the same way as Markus (2005), in saying that English uses pitch for “sentence stress, whose domain is not the sentence but the tone group” (p. 989). Jenkins (2000) proposes a global model of English, where nucleus-stress is defined as the most important feature for intelligibility, compared with pitch movement and tone units. Jenkins (2000) recommends teaching nuclear-stress placement rules in EFL classrooms for effective English communication.

Markus (2005) describes how pitch, length, and loudness are variously combined and used for English stress patterns. With regard to the supra-segmental stress patterns, English marks the prominence of the tone-unit nucleus. Nucleus has prominence with pitch, loudness, and length combined, and in other parts of the tone group, English relies more on a functional pitch contour (Markus, 2005, p. 111). Therefore, prosodic features such as pitch, loudness, and length are used in various combinations to function as word stress, tone group nucleus, or tone group stress: These are listed in Table 1 below.

Table 1

*Functions of Prosodic Features in English and German (Markus, 2005, p. 111)*

| Unit | English | German |
|---|---|---|
| Word stress | length, loudness, (pitch) | pitch, (loudness) |
| Tone group nucleus | pitch, loudness, length | pitch, loudness, length |
| Tone group stress | pitch | length, loudness, pitch |

These tone group nuclei appear at regular intervals, which creates rhythm in English sentences.

Table 2

*Different Prosodic Types in a Discourse (Markus, 2005, p.116)*

| Pre-head | Head | Body | Nucleus | Tail |
|---|---|---|---|---|
| I have studied | ver- | y hard for | All | the subjects. |
| It was | pleas- | Ant | talk- | ing with you. |
| We | hope | you will get | well. | |
| | Thank | You | ver- | y much. |
| | | His um- | brel- | la! |
| | | Good | morn- | ing. |
| | | | Hi! | |

According to Markus (2005), the regularity of English rhythm is due to the recurrence of stressed syllables. Markus (2005) notes that in English, "stressed syllables tend to occur at approximately regular intervals so that the rhythmical units known as feet[6] can be constructed." In each foot, "unstressed syllables tend to be compressed and the stressed syllables tend to be lengthened" (p. 109). English is a good example of a stress-timed language and tends to adapt the quantity of syllables to the rhythm conditioned by word stress, thus producing units that are similar to metrical feet. Therefore, English given syllables are subject to its rhythm and are often unpronounced or prolonged, to adjust to the rhythm. German also has a “foot-timed” prosodic system and, therefore, it is no less stress-timed than English is. However, German does not have the same vowel reduction as English so that its vowels retain more of their quality.

With its unique rhythmic pattern, English is described as a “rhythmic language” for good reason. Adding to this supra-segmental phonological aspect of prosody, the functional aspect of prosody will be described, referring to Ichikawa (2011), in the next section.

**2.1.6 Ichikawa’s definition of linguistic prosody as preliminary announcement information.** One aspect of prosody is intonation, which conveys information in speech. Ichikawa (2011) calls such information *prosody information*. The results of his speech prosody analysis suggest that prediction using such information as pitch and chunking, or so-called *preliminary announcement information*, which is understood through speech prosody, can achieve the smooth progression of a dialog in real time (see Table 3).

In linguistics, the information conveyed in spoken language has been categorized into three kinds of information: Linguistic information (lexical, syntactic, semantic, and pragmatic information), paralinguistic information (intentional, attitudinal, and stylistic information), and nonlinguistic information (physical and emotional information) (Fujisaki, 1996). Comparatively, Ichikawa proposes that information in real-time dialogs can be divided into two major categories: “message information (linguistic information and utterance person information) and information supporting real-time transition (message preliminary announcement information and speaker-shift preliminary announcement information)” (pp. 235–236). The latter is the information added at the time an utterance is expressed. This *message preliminary announcement information* can be observed in a change in the rhythm of a dialog caused by the condition of the partner through speech prosody.

Table 3

*Structure of Real-Time Dialogue (Ichikawa, 2011, p. 179)*

| Information conveyed through real-time dialogs | | | | Physical characteristics |
|---|---|---|---|---|
| Message information | Linguistics related information | Linguistic information | Phonological information without prosody such as intonation | Spectrum |
| | | | Paralinguistic information | Prosody |
| | | Sentence structure information | Segmentation of words and accented phrases, i.e., chunking and focus | |
| | | | Modification structure (structural relationship among words in a sentence) | |
| | Utterance person information | | Individual nature (sex, age, individuality) | |
| | | | Emotion information | |
| | | | Physical health | |
| | | | Social relationship | |
| Information supporting real-time transition | Message preliminary announcement information, conveyed with phonology, vocabulary, sentence structures | | Phonology preliminary announcement information | Prosody |
| | | | Segmentation preliminary announcement information, conveyed by words and accented phrases | Prosody |
| | | | Preliminary modified structure | |
| | Speaker-shift preliminary announcement information known by speakers' attitudes | | | |

Table 3 shows that information is conveyed in real-time dialogs through its physical characteristics, i.e., spectrum or prosody. In Ichikawa's chart, spectrum and prosody share the role of information messenger, whereas prosody alone has an additional important function as the carrier of information that supports real-time transitions (i.e., message preliminary announcement information).

Markus and Ichikawa each provided insight into English prosody. Markus (2005) provided an explanation of the phonological characteristics of English prosody, and to these, Ichikawa (2011) added the communicative functions of prosody. Other research on prosody perception/production in English will be reviewed in the next section.

**2.1.7 Previous studies on language prosody processing in EFL.** Generally, the term prosody is used to express prosodic features such as rhythm and tone in language or sometimes in music. In academic research, music prosody has been studied by Jusczyk et al. (2001), while linguistic prosody has been studied by Patel and Daniele (2003).

Non-native speakers' language prosody has been researched in Juffs (1990) and Wennerstrom (2001) as follows: Juffs (1990) measured word stress and sentence-level pitch accent in an oral reading task by Chinese speakers of English. He found that errors were mostly due to the placement of word stress, i.e., heavy stress on every word. Juffs's study suggests that word stress is acquired in a unique way by Chinese speakers of English at intermediate stages of prosody acquisition. This *intermediate prosody* reveals patterns that are characteristic of neither the native nor target languages (Wennerstrom, 2001, p. 248). It seems worthwhile, therefore, to study intermediate prosody in comparison to native and nonnative prosody.

In the following studies, nonnative speaker prosody is defined as the effect of instruction on acquisition. In Gilbert's (2008) study, the participants listened to filtered English utterances in which only the prosody was accessible and then mimicked the intonation using the musical toys "kazoos," in order to improve their pronunciation. Gilbert's (2008) study introduced his successful educational practices and recommended teaching prosody in EFL. Similarly, in Neufeld's (1978) study, native English speaking students at Ottawa University practiced prosodic articulatory features of two non-Indo-European languages without knowing the meaning of what they were uttering. The program was 18 hours long and learners were encouraged to listen attentively, sensitize and trace intonational contours, distinguish between phonemes in minimal pair contexts, and finally, imitate the devoiced or whispered utterances orally. The 20 participants were selected for their high language aptitude scores: Native speakers of the two languages judged the utterances imitated by the participants. After the utterances were scored on the 1-5 point scale, 8 to 9 out of 20 participants were judged to be native speakers of each language in tests. The consistency of their performances in both languages ($r$ = 0.82) was proved. This study demonstrates that young adults can acquire native or near native proficiency by learning the sound patterns of new languages without systematic or semantic knowledge. There was no significant difference in subjective learning difficulty between the two languages nor any significant influence of the testing order.

Both studies found positive results in initiating language acquisition from oral and aural intonation practice. These studies suggest that focusing on the prosody of a target language at the initial or early stages of acquisition could help learners to acquire native-like speech.

When we listen to the prosodic flow of an utterance recursively, it will come to sound familiar to us and somehow support our comprehension. Prosody functions in expressing affection, and in creating intonation, chunking, and focusing in a sentence (Peppé, McCann, & Gibbon, 2007). Therefore, it should be worth investigating how to cultivate language prosody perception and production.

From the previous research described in this section, processing of prosodic elements of language might be a key factor in the auditory comprehension and oral production of language. In the next section, prosody perception and production will be discussed in terms of enhancing communicative proficiency.

#### 2.1.8 Correlation between English prosody processing and English pronunciation.

The phonological structure and communicative function of English prosody have been described in the previous sections. Considering the phonetic and semantic information that prosody carries, prosody perception/production will help learners to communicate smoothly. The following research studies imply that those who process English prosody well can be expected to perceive and perform well in English communication.

Anderson-Hsieh et al. (1992) analyzed 60 oral reading passages by male nonnative speakers of English at a range of proficiency levels and from a variety of language backgrounds. They rated the pronunciation of each speaker on a 7-point scale and then analyzed the speech for accuracy in three categories: segments, syllable structure, and prosody (including stress, rhythm, phrasing, and intonation). The researchers found, from the standardized regression analysis, that prosody had a greater influence on the pronunciation rating than the influence from the segmental or syllable structure error rates. The standardized multiple regression analysis revealed an $R^2$ of 0.89 and the highest coefficients[7] with prosody ($\beta = 0.58$, $p < .001$), the second highest coefficients with segmental error rate ($\beta = -0.27$, $p < .001$), and the third highest coefficients with syllable structure error rate ($\beta = -0.25$, $p < .001$).

English is a stress-timed language, in which stressed vowels appear at certain intervals and create a rhythm (see the review of Markus (2005) above). Thus, English has a rhythmic flow when spoken, and it has a prosodic character of nucleus-stress placement. Thus, better command of the nucleus-stress placement may correlate with better pronunciation of English. Considering the rhythmic character of English prosody, the rhythm processing of language will be discussed, in comparison with that of music, in the next section.

**2.1.9 Correlation between musical and linguistic rhythm processing.** Making much of the importance of prosody perception/production in English communication, further research will be meaningful to ascertain what aptitude contributes to English prosody processing skills and English proficiency for native Japanese children. In the previous studies described in this section, the correlation between pitch and rhythm perception in music and language was investigated and rather positive results were obtained. In order to improve the sensitivity and productivity of language rhythm and pitch, early music training might implicitly have a positive effect, since one of the language aptitudes is considered to be audio perceptual acuity (Doughty, 2013).

Apparently, stress-timed English and mora-timed Japanese have quite different prosodies, as may English classical instrumental music and Japanese classical instrumental music, according to Patel and Danielle (2003)'s suggestion that composers' language rhythms may influence the rhythm of their musical works.

Jacobsen and Imhoof (1974) found that music training is not necessarily a factor for succeeding in learning Japanese as a second language, in spite of a number of studies that suggest the effectiveness of music training on pitch detection sensitivity for French (Schön et al., 2004; Magne et al., 2006), Portuguese (Marques et al., 2007), and Spanish (Thompson et al., 2004). Jacobsen and Imhoof (1974) suggested that pitch discrimination might not be a crucial factor for the acquisition of Japanese, but may contribute to acquiring more melodious and rhythmical language.

The correlation between language and music rhythms has also been empirically studied by Patel and Daniele (2003), whose results supported the idea that the prosody of a composer's native language can influence the rhythm structure of his or her musical works. They compared the rhythm in music and language of 16 British composers, including Bax, Elgar, Holst, Ireland, et al., and the French composers Debussy, Honegger, Ravel, Roussel, et al. They used the rhythmic index, normalized pairwise variability index[8] (nPVI), which is a "purely relative measure of variability" (Patel & Daniele, 2003, B37). It represents the durational difference between each pair of intervals as measured relative to the average length of the pair. This rhythmic standard of languages was conceptualized by Low and Grabe (1995). Low and Grabe (2002) measured vowel duration and the duration of intervals between vowels in a passage of speech, instead of measuring syllable duration alone.

Patel and Daniele (2003) found that English and French music were significantly different in terms of their nPVI values, as were spoken English and French, although the observed difference in music nPVI values was smaller than that for speech. In their experiment, the rhythmic index, nPVI, of British English was found to be larger than that of French; similarly, the British musical nPVI value was found to be larger than that of French music. They successfully measured the musical nPVIs using composers' notations—in this study, those of Debussy and Elgar. The relative duration of each note was shown below the musical notes (p. B40). As shown in the figure (Patel & Daniele, 2003, p. B41), the results of musical nPVI measurements were similar to the results of linguistic nPVI measurements (p. 38). The average nPVI values for British English and French musical themes turned out to be different: British English music had a greater value than French music (mean nPVI = 46.91 vs. 40.90) and the difference was found to be significant (Mann-Whitney *U*-test, $U = 9993.5$, $p < .01$).

Patel and Daniele (2003) concluded that the discrepancy between the rhythm structures of British English and French music might be due to the characteristics of the composers' native languages. Their study suggested that a composer's language prosody characterizes his/her music prosody, which means that composers with different mother tongues have created the musical pieces with different prosodies. It is insightful that composers' language rhythms may influence the rhythm of their musical works.

The rhythmic difference between Western and Japanese music or poetry can be seen in an anacrusis, or auftakt in German. Western musical compositions or poetry often begin with an anacrusis, whereas Japanese ones never do. The difference in language rhythms, calculated with nPVI, comes from the two different rhythmic structures where the "stress-timed rhythm of English allows vowel reduction, in contrast with syllable-timed rhythm of French, which does not" (Ramus, 2002b, p. 1). In Ramus's study, the duration and variability of "inter-vocalic intervals" and "vocalic intervals" were compared between eight languages, considering the phenomenon that stress-timed languages tend to allow more complex syllables, and longer and more variable sequences of consonants than syllable-timed languages. *Inter-vocalic*, i.e., *consonantal, intervals* are made up of consonants and sequences of successive consonants, while *vocalic intervals* are vowels and sequences of consecutive vowels (Ramus, 2002b, p. 1). Each difference between the inter-vocalic intervals and vocalic intervals was normalized by their average duration in nPVI. From the nPVI figure (Ramus, 2002b, p. 1), languages were clustered in groups that resemble rhythm classes: English, Dutch, and Polish as stress-timed languages, French, Spanish, Italian, and Catalan as syllable-timed languages, and Japanese as a mora-timed language. The figure (Ramus, 2002b, p.1) illustrated the *rhythm class hypothesis*.

According to the studies on stress-timed languages by Grabe (2002) and Satoi (2003), the intervals between stresses or rhythmic feet are of near-equal length, whereas in syllable-timed languages, successive syllables are of near-equal length. The third type of rhythm, mora-timing, is exemplified by Japanese. In mora-timing, successive mora are near equal in duration. These rhythmic differences among languages make foreign language acquisition challenging. Using nPVI values, Ramus (2002b) and Warner and Arai (2001) compared the rhythms of mora-timed, stress-timed, and syllable-timed languages, and discussed the rhythm typology from the aspect of advocates and opponents and the need for further research with larger corpora.

Just as composers of different languages tend to compose using different rhythm structures, the rhythm of songs, dances, and instrumental music also varies, although songs are not directly targeted nor studied in this thesis. The Japanese national anthem, bon-dances, and Tsugaru-shamisen all reflect the mora-timed rhythm of the Japanese language (Ichikawa, 2007). Just as Japanese classical music reflects Japanese prosody, European classical instrumental music was empirically found to be affected by the prosody of a composer’s native language (Patel & Daniele, 2003). Music training as implicit prosody training in the Japanese EFL environment will be proposed in the next section.

**2.1.10 Proposal to use music training as implicit prosody training in the Japanese EFL environment.** Western and American music is generally perceived as being more rhythmic than Japanese music, and requires rhythmic sensitivity to perform. Although Japanese children have become familiarized with Western and American music these days, they continue to have difficulty communicating with the appropriate English prosody. Nakano and Natsume (2011) explained that the difficulty faced by Japanese EFL learners in using English prosody derives from the difference between their accustomed mora-timed rhythm and the English stress-timed rhythm. Several studies have been conducted on the relationship between rhythm processing ability and Japanese children's English prosodic ability (Nakano, 2012). Nakano and Natsume's (2011) empirical study suggested that the beat in the rhythm instruction material (RIM) developed as original material in their study might enhance English rhythm learning and acquisition. Nakano and Natsume (2011) reported that the frontal lobes of the brain were bilaterally activated when the participants attempted the RIM in English, in contrast to the generalization that language processing activates left temporal sites.

Aoyama and Guion (2007) found from the exploratory analyses of duration and F0 range that there were measurable prosodic differences between the English spoken by L2 Japanese speakers and native English speakers, both in duration and range, and that function words were pronounced proportionally longer by L2 Japanese speakers. These studies suggest that each language has its own rhythmic pattern and pitch contour pattern so that rhythm and tone perception/production may be crucial for achieving native-like fluency in a target language.

In previous research, English prosody has been studied as a branch of linguistics or communication studies, and more discussion is required on how educators can promote prosody processing and production of English in classrooms in Japan. For school-aged learners to become familiar with English prosody in Japanese EFL environments, what aptitude should they possess? Sound sensitivity, i.e., rhythm sensitivity and tone sensitivity, may help those learners to process English prosody, based on the implications from previous research on the correlation between sound sensitivity and English prosody processing. This leads to the further implication that early training in Western classical musical instruments helps EFL learners to enhance sound sensitivity for English. Considering this idea of implicit English prosody learning, the previous studies on (1) the relationship between sensitivity to musical and linguistic sounds, and (2) the relationship between linguistic prosody processing skill and overall linguistic skill will be reviewed, from the next section.

## 2.2 Relationship between Sound Sensitivity and L1 Prosody Processing Skill

A number of researchers have proposed that music training affects trainees' pitch processing in language and music (Schön et al., 2004), episodic memory in music and potentially in speech (Jusczyk et al., 2001), and emotional prosodic sensitivity (Thompson et al., 2004). Thompson et al. (2004) reveal that music lessons promote sensitivity to the emotions conveyed by speech prosody. In this section, the relevant research on musical sound sensitivity and native language prosody processing skill will be reviewed. In the previous studies concerning music learning experience, the following requirements for music training were applied to musically trained participants:

- There were seven-year-olds who had completed one year of formal music training for 45 minutes and home practice for 10–15 minutes a week (Thompson et al., 2004). The musically trained participants showed a significant advantage in extracting prosodic information in speech.
- There were eight-year-old participants with 4±1 years of music training on average (Magne et al., 2006). The musically trained participants outperformed the untrained ones in detecting a weak incongruity in both music and language.
- There were musically trained participants with more than 15 years of musical lessons

(Thompson et al., 2003). The musically trained participants outperformed the untrained ones at identifying various emotions.

- There were four groups of 12 participants (age range = 15–50 years) composed of two musician groups from the regional musical school with at least seven years of training in classical music (Besson & Faïta, 1995). The musically trained children outperformed the untrained ones in the recognition of familiar musical melodies and in classifying terminal notes.
- There were 20 eight-year-old children with general music training of eight weeks (twice a week for 50 minutes) or six months (twice a week for 75 minutes) (Besson, Magne, Moreno, Schön, & Santos, 2007). The results demonstrated an increased auditory acuity in the musically trained participants compared to the untrained ones, and transfer of training effects between music and speech, as revealed by a lower error rate of the musically trained participants compared to the untrained ones with regard to the weak incongruity in speech (p. 404).
- There were participants with 4–13 years of private instruction on a musical instrument (mean age = 8.1 years) (Jusczyk et al., 2001). The musically trained children could more accurately identify excerpts whose prosodic cues conflicted with the structure of the melodic context, in

the familiarized excerpts placed in various melodic contexts.

- There were nine musically trained participants (mean age = 31 years) with 15 years of music training on average (Schön et al., 2004). The result will be discussed later in this chapter.

As already defined in Chapter 1, the age of 9 years is the critical cut-off point for music acquisition, as described in Gordon (1982). Overviewing these requirements for being regarded as musically trained, music training will be discussed in relation to language prosody processing, in numerous empirical situations. Previous studies on the relationship between sound sensitivity and first language (L1) prosody processing skill are reviewed in sections 2.2.1 to 2.2.3.

**2.2.1. Effectiveness of music training in facilitating L1 pitch processing in adults.**

Schön et al. (2004) and Magne et al. (2006) have suggested that intensive music training affects the recognition of pitch contours in a language. They concluded that music training facilitates pitch processing in both music and language.

The experimental study of Schön et al. (2004) was conducted with participants consisting of 18 French native speakers aged 31 years, including nine musicians. In the test, the participants were told to listen attentively and identify the intonation of 120 French sentences, i.e., their prosodies and melodies, which consisted of 40 strongly incongruous, 40 weakly incongruous, and 40 congruous prosodies. In these test sentences, the F0 (fundamental frequency) of the final word was increased using the software Winpitch (Martin, 1996) by 35% for weak incongruity, and by 120% for strong incongruity. In the test notes, the F0 of the final note was increased by one fifth of a tone for weak incongruity, and by a halftone for strong incongruity.

With regard to pitch detection, the musicians scored significantly better in distinguishing incongruous language prosodies and melodies, i.e., detection of pitch violations in both music and language, than non-musicians. More importantly, musicians detected weak incongruity significantly better than non-musicians did, not only in music but also in language. Through their experiment, it emerged that the musicians detected weak F0 manipulations better than non-musicians did, with overall shorter onset latency for musicians than for non-musicians. Unfortunately, Schön et al. (2004) did not comment on the relatedness of the degree of pitch manipulation in language to the degree of tone manipulation in music. In the case of language, the pitch manipulation for strong incongruity was more than one octave (135%). In the case of music, the manipulation for strong incongruity was only a semitone. The manipulation for language was thus more than 10 times as strong as that for music (one octave corresponds to twelve semitones), which is remarkable. It is very likely that humans are much more sensitive in music to tonal incongruities, than in spoken language to pitch incongruities of the same extent. If so, it would be quite understandable that musical training would also help with pitch processing in language. However, the opposite—a positive effect of language training (including familiarity with intonation and pitch) on tonal processing in music—would not follow easily from these considerations on intervals and sensitivity.

In the experimental study of Schön et al. (2004), there was also a difference in the brain area that reacted to the sounds between those two groups. The scalp distribution of the congruity/incongruity effect differed between musicians and non-musicians. Whereas for musicians, the negativity was largest over temporal sites bilaterally, for non-musicians, it was largest over left temporal sites (p. 346). This brain reaction study will be reviewed in more detail later in this thesis.

In Thompson et al. (2003), the musically trained adults were found to be statistically superior in their ability to match spoken utterances with their tonal analogues (only the tone sequences that retained pitch and temporal patterns) than untrained adults. The trained adults judged more correctly whether the intonation melody matched the prosody of the phrase or not. This study indicates that musically trained adults have an enhanced ability to extract prosodic information from spoken phrases.

#### 2.2.2 Effectiveness of music training in facilitating L1 pitch processing in children.

In the same type of Event-Related Potential (ERP) study as that conducted by Schön et al. (2004), Magne et al. (2006) conducted an equivalent empirical study with child participants: Twenty-six children (14 girls and 12 boys; age 8 ±1 years), 13 of whom were musicians and 13 non-musicians, participated in the experiment, which lasted for approximately two hours. The musician children possessed 4±1 years of music training on average. All children were native speakers of French. All musician children played an instrument (violin = 5, guitar = 2, flute = 1, clarinet = 2, harp = 1, piano = 2), which they regularly practiced every day for approximately 20 to 30 minutes. They also took music lessons twice a week for half an hour. Thus, these children played music for approximately three–four hours per week. Magne et al. (2006) suggested the existence of neurophysiological processes, which may underlie positive transfer effects between music and their native language in children. Magne et al. (2006) researched the reaction of children to pitch processing in music and language. The percentage of errors in the detection of weak pitch incongruity was lower for musicians than for non-musician children not only in music but also in language. ERP data revealed that while a negative-positive complex (N200: negative in 200msec–P600: positive 600msec) was elicited by strong incongruities both in the musician and non-musician groups, a negative component (N300: negative in 300msec) developed in response to weak incongruities only in the musician group. The result demonstrated that increased auditory acuity has an electrophysiological correlate in the occurrence of the N300 component only in musicians. In addition, the early negative reaction of musician children to the weak incongruity in music may reflect a greater sensitivity to pitch.

Comparison of these results with the results obtained from adults (Schön et al., 2004) suggests that both adult and child musicians detected weak incongruities in their first language better than non-musicians did. Additionally, both studies implied a positive transfer effect between two cognitive domains in the brain in the musicians: the music and language domains.

**2.2.3 Relationship between sound sensitivity and L1 phonological awareness.** The former articles dealt with the relationship between L1 prosody and musical pitch contour. Articles on L1 phonological awareness and sound sensitivity will be discussed in this section.

Peynircioğlu (2002) conducted an experimental study with 32 Turkish and 40 English native speaker participants of four–six years old. The participants took the rhythm and pitch subtests from the Seashore Measures of Music Talents (Seashore, Lewis, & Saaeteweit, 1960), the phoneme deletion task[9] with words in their native languages and with pseudo-words, and the tone deletion task. Thirty-two Turkish children took the phoneme deletion task including 10 initial vowel deletion tasks and 10 initial consonant deletion tasks. Secondarily, 10 final vowel deletion tasks and 10 final consonant deletion tasks were administered to the participants. The participants were grouped into high or low sound sensitivity groups, based on their scores in the

Seashore test. Probably because the characteristics of the Turkish language with frequent final vowel manipulation had an influence, the final phoneme deletion scores were higher than the initial phoneme deletion scores in both high and low sound sensitivity groups. High sound sensitivity children did significantly better in the majority of the phoneme deletion tasks. Furthermore, the participants in both groups did better on final tone deletion ($F(1,30) = 25.82$, $MSe = 8.71$, $p < .001$), and high sound sensitivity children did significantly better in the final tone deletion task ($F(1,30) = 34.46$, $MSe = 20.00$, $p < .001$).

Forty English-speaking children performed more accurately in the consonant deletion tasks than the vowel tasks. Although final phonemes were manipulated more successfully than initial phonemes by both groups ($F(1, 38) = 3.97$, $MSe = 1.89$, $p < .005$), high sound sensitivity children did significantly better overall in the phoneme deletion tasks ($F(1,38) = 52.53$, $MSe = 5.20$, $p < .001$). The participants in both groups did better in final tone manipulation, and high sound sensitivity children did significantly better in all tone deletion tasks ($F(1,38) = 61.61$, $MSe = 2.16$, $p < .001$). For the English-speaking children, both initial and final tone manipulations were highly correlated with phonological awareness. In both Turkish and English groups, high sound sensitivity children performed better overall in the tasks. As a result, high sound sensitivity children in both languages emerged as sensitive to musical and linguistic sounds and as having better phonological awareness.

However, the phonological unit in this study was not supra-segmental, but in segmental units. Sentence prosody and phonemes work as constituents of sentences at different levels. A phoneme is linguistically defined as the minimal unit of a word sound, and prosody is a characteristic melody or rhythm or syntax of the sentences of any language, i.e., a unit in supra-segmental phonology. Thus, the relationship between supra-segmental phonology and sound sensitivity should be further researched.

**2.2.4 Music training and L1 verbal memory.** Research has been conducted on the correlation between sound sensitivity and verbal memory. Chan, Ho, and Cheung (2003) compared 45 musically trained male children aged 6–15 (with at least one hour of traditional classical music lessons in either violin or flute per week from professional instructors at the Hong Kong Academy for Performing Arts for one–five years) with 45 male children aged 6–15 who did not have such music training. They found that musically trained children performed significantly better than non-musicians at verbal memory tasks. Verbal memory was measured with HKLLT-Fork One, which is a sensitive verbal memory test (Chan et al., 2003). The test comprises a list of 16 two-character Chinese words, which was presented orally to each participant three times. The participants were asked to recall as many words as possible in the three learning trials and 10-min and 30-min delayed recall trials. Their study indicated that music training systematically affects memory processing: the Pearson product-moment correlation on the total verbal learning score and the duration of music training yielded a significant result ($r = .59$, $p < .001$). Non-musical factors that conflicted with this relation were examined and it was found that total verbal learning score was significantly correlated with age ($r = .50$, $p < .001$) and educational level ($r = .51$, $p < .001$). From this study, regular music training may potentially improve one's verbal memory, although verbal memory depends partially on educational level and age.

**2.2.5 Music training and L1 speech segmentation skill.** François and Schön (2014) conducted research on the effectiveness of musical practice for enhancing statistical regularities in speech. As already referred to briefly in Chapter 1, they proved the existence of enhanced neural sensitivity to statistical regularities in musicians. They explained that infants have the same ability to discriminate speech sounds, and retain sensitivity to statistical regularities presented in their mother language. However, non-musicians mostly lose this while musicians retain this sensitivity throughout life. They also support the hypothesis of positive transfer of training effect from music to sound stream segmentation, i.e., speech segmentation. They proposed that musical practice works effectively on brain functions to strengthen auditory segmentation skills. Thus, they propose the idea of promoting the effective use of music “as a tool for optimizing speech acquisition trajectories both in normal and impaired populations” (p. 127).

François and Schön (2011) used artificial sung language learning coupled with the electrophysiological approach. ERP data showed that musicians learned better than non-musicians in both the musical and linguistic structures of the sung language. In their study, the participants comprised 16 professional musicians (mean age 27) with more than 12 years of formal musical learning and three–seven hours of daily practice, and 20 non-musicians (mean age 25) with less than two years of formal music training. In the learning phase, the participants were instructed to listen to tri-syllabic sung words such as gimysy C3 D3 F3, mimosi E4 Db4 G3, pogysi D4 C3 G3, pymiso B3 E4 F4, and sipygy G3 B3 C4. In the Linguistic test, the participants had to listen and choose the word that closely resembled what they had just heard in the learning phase from each of 25 pairs of word stimuli, by pressing one or two buttons on a computer keyboard. In the Linguistic test, the five unfamiliar stimuli comprised gysimi, mosigi, pygymi, sogimy, and sypogy. In the Musical test, the participants had to listen to a piano melody and choose which of two melody stimuli most closely resembled what they had just heard in the learning phase. In the Musical test, the unfamiliar stimuli were made up of the same melodic contour as the unfamiliar words stimuli. Although not significant, there was a tendency for musicians to have a higher level of performance than non-musicians in both tests. ERP data revealed the musical training effects on the N1 (initial negative) component with a larger amplitude for musicians than non-musicians, in listening to the unfamiliar music and linguistic test stimuli. This result suggests that musicians pay more attention to general sounds and even more selective attention to musical stimuli than non-musicians.

Marques et al. (2007) have proposed that music training affects trainees' pitch processing. There also exists a relevant study on music performance and episodic memory, based upon the theory of speech perception aided by prosodic memories (Jusczyk et al., 2001). Slevc and Miyake (2006) dealt with phonological awareness in supra-segmental phonological units, in relation to musical tonal memory.

**2.2.6 Sound sensitivity, L1 phonological awareness, and reading proficiency.** Anvari et al. (2002) indicate that sound processing skill in rhythm, melody, and chords is reliably related to phonological awareness and reading. They examined the relationships between phonological awareness, music perception skills, and early reading skills in a population of 100 four- and five-year-old children. The materials used for measuring phonemic awareness were Rhyme Generation, Blending, Oddity, and the Rosner Test of Auditory Analytic Skills. The material used for reading was the Wide Range Achievement Test-3 (WRAT). The materials used for measuring music were Same/Different Rhyme Discrimination, Same/Different Melody Discrimination, Same/Different Chord Discrimination, Rhythm Production, and Chord Analysis. They found that music skills correlated significantly with both phonological awareness ($r$ = .59, $p$ < .01) and reading ($r$ = .57, $p$ < .01) in four-year-old children, while in five-year-old children, pitch processing skill correlated only with reading ($r$ = .45, $p$ < .01). Hierarchical regression analysis showed that phonemic awareness and music skills predicted reading ($R^2$ = .39, $p$ < .009) in five-year-old children, whereas phonemic awareness and pitch predicted reading ($R^2$ = .31, $p$ < .017) in four-year-old children.

Several other studies have demonstrated the relationship between music training and L1 reading ability. For example, Atterbury (1985) found that poor readers aged seven to nine years in the United States were significantly impaired in melodic/harmonic tone discrimination and rhythm production. Douglas and Willats (1994) concluded that rhythm discrimination correlated with reading ability in seven- and eight-year-old children. Lamb and Gregory (1993) have found that four- to five-year-old children who were good at melodic/harmonic pitch discrimination also did well on phonemic awareness and were good readers. The outcomes of these research studies imply that some musical skills can predict aspects of L1 verbal ability, such as verbal memory, phonemic awareness, and reading ability.

The correlation and some causal relationships between phonemic awareness, musical skills, and reading skills in these children indicate the importance of paying attention to phonemes in language acquisition and pitch in early musical education. Phonemic awareness tested in phonemic discrimination and categorization has been emphasized in FLA studies (Doughty et al, 2010), and the present study aims to draw foreign language teachers' and researchers' attention to the potential benefit of prosody processing development in the language acquisition process.

## 2.3 Relationship between Sound Sensitivity and L2 Phonological Processing Ability

Considering the studies on L1 prosody processing ability introduced above, previous studies on the relationship between sound processing skill and L2 phonology are reviewed in sections 2.3.1 to 2.3.3.

**2.3.1 Relevance of sound sensitivity to L2 receptive/productive phonology.** Slevc and Miyake (2006) posed the question whether sound sensitivity affects individual differences in English receptive/productive phonology. Fifty native speakers of Japanese aged from 19 to 52 years participated in their study, none of whom were immersed in English until after the age of

11, so-called late learners, and all had lived in the United States for more than six months at the time of testing.

The researchers administered the Receptive phonology tests at the word level, where the participants chose the pronounced word from 26 minimal pairs, such as “clown” and “crown.” Secondly, the Phonology test used at the sentence level was a choice test, where the participants listened to 26 minimal sentence pairs and chose the one that was meaningful: for example, “Some researchers believe that playing/praying is an important part of mental development” (p. 676). Thirdly, at the passage level, the participants listened to passages and chose 21 mispronounced words out of 43 words underlined in a written transcript of the recording.

In the Productive phonology test, the participants read the same sentences as given for the Receptive phonology test. The native speakers checked the phonology, pronunciation, intelligibility, and prosody on a 9-point scale. As another phonological skill, they measured phonological short-term memory with a repetition task of nonwords. Each repetition task was one–four syllables long. As a measurement of sound sensitivity, the Chord analysis subtest, the Pitch change subtest, and the Tonal memory subtest in the Wing Measures of Musical Talents (Wing, 1968) were conducted.

Statistical correlations were found between sound sensitivity and receptive/productive phonology ($r = .52$, $p < .005$; $r = .45$, $p < .005$). Furthermore, as a result of the hierarchical regression analysis, sound sensitivity significantly accounted for L2 receptive phonology and productive phonology ($\Delta R^2 = .12$, $\beta = .37$, $p < .005$; $\Delta R^2 = .08$, $\beta = .30$, $p < .05$). This study suggests a correlation between L2 phonological perception/production and sound sensitivity. The next study shows a correlation between tonal memory and L2 phonological tone identification.

**2.3.2 Effectiveness of music training for L2 phonological tone identification.** The effect of music training on learning L2 speech was contrastively studied in Gottfried (2007). Thirty-eight nonnative speakers of Mandarin participated in his study, of whom 24 were music majors, and the other 14 had less than five years of music training. In serial experiments, the participants attempted a tone glide identification task[10] and an Imitation test of Mandarin syllables. The third task, a Mandarin tone identification task,[11] included the intact and silent-center identifications.

In the results, music majors scored significantly higher in the tone glide identification task and Mandarin silent-center identification task. As a result of correlation analysis, years of musical study and tone glide identification score were significantly correlated ($r = .65$, $p < .01$), as were tone glide identification and Mandarin silent-center identification scores ($r = .47$, $p < .05$). In this study, Gottfried (2007) suggested that music training might affect phonological tone identification in tone languages such as Mandarin Chinese.

In Marie, Delogu, Lampis, Belardinelli, and Besson (2011), a same-different task was used to test the hypothesis that those with musical expertise perform better in the discrimination of tonal and segmental (consonant, vowel) variations in a tone language, Mandarin Chinese, than non-musicians. The participants comprised 11 musicians (mean age = 24 (*SD* = 2.6) years, age range = 21–29 years, seven female) and 11 non-musicians (mean age = 23 (*SD* = 1.9) years, age range = 21–28 years, six female). All were native speakers of French and had no experience of tone languages. Pairs of four-word sequences (identical or different prime-target four-word sequences) were presented to them in order for them to decide whether the sequences were the same or different by pressing a key quickly and accurately in response; event-related brain potentials recorded their reactions. Two blocks of 42 pairs were presented for segmental variations and another two blocks of 42 pairs were presented for tonal variations. The musicians detected both tonal and segmental variations more accurately than non-musicians did. Moreover, tonal variations were associated with higher error rates than segmental variations and elicited an increased N2/N3 (negative 2/negative 3) component that developed 100 milliseconds earlier in musicians than in non-musicians. This study implies that musical expertise may have higher sensitivity to basic acoustic parameters such as is also necessary in foreign language learning.

**2.3.3 Sound sensitivity and L2 pronunciation skill.** Milovanov et al. (2008) found a correlation between English pronunciation skill and sound sensitivity, including pitch discrimination ability, sense of rhythm, and sense of tonality, in 10-12 year-old participants learning English as a second language in Finland. To ensure that all participants had an adequate amount of pre-training in English pronunciation, they received an eight-week course in English pronunciation training, including phonemic discrimination exercises. After the pronunciation training, they were grouped into an advanced pronunciation group (mean accuracy scores: 74%) and less-advanced group (mean accuracy scores: 46%). Their musical ability was measured with the Seashore subtests, and the one-way ANOVA indicated that the advanced pronunciation group were superior to the less-advanced group in terms of pitch discrimination ability ($F(1,38) = 21.27$, $p < .001$), timbre ($F(1,38) = 5.36$, $p < .05$), sense of rhythm ($F(1,38) = 6.63$, $p < .001$), and sense of tonality ($F(1,38) = 6.71$, $p < .001$). They suggested that sound sensitivity and linguistic skills could partly be based on shared neural mechanisms, and that musical sound sensitivity and linguistic skills are interconnected. Concretely, their ERP recordings and behavioral measures revealed the sound processing accuracy of children correlated with accurate pronunciation skills.

Milovanov et al. (2009) examined children’s preattentive duration processing ability in language and music stimuli. Forty Finnish child volunteers aged 10–12 were grouped into either advanced or less-advanced foreign language production skill groups, and into either advanced or less-advanced musical aptitude groups. The advanced pronunciation and musical skill group displayed stronger frontal mismatch negativities (MMN) to duration changes in both speech and musical sounds. In ERPs, the music stimulus was a violin tone (tone C-major sound: C4, fundamental frequency 261.3 Hz). In ERPs, the standard and deviant duration of the speech stimuli were 250 ms (milliseconds) and 150 ms, respectively, which were equivalent to the speech sounds. The standard speech stimulus consisted of a 250 ms monaural Finnish vowel /ö/. The deviant speech stimulus /ö/ was otherwise similar but only 150 ms in duration.

In the vowel condition in the speech stimulus, the MMN to durational changes in the deviant and standard stimulus was statistically significant for the advanced English pronunciation group ($t(59) = -6.39$, $p < .001$) and also for the less advanced group ($t(59) = -4.0$, $p < .001$). In the violin sound condition, a statistically significant MMN to the durational changes in the deviant and standard stimuli was obtained for both the less-advanced pronunciation group ($t(59) = -7.67$, $p < .001$) and the advanced pronunciation group ($t(59) = -7.46$, $p < .001$). Unexpectedly, this result did not find a significant difference in MMN between the advanced and less-advanced pronunciation groups. However, in ERPs, the advanced pronunciation group demonstrated stronger frontal responses than the less-advanced group in both the vowel and violin conditions ($F$ (2, 236) = 4.39, $p$ < .005). Moreover, both groups had stronger responses to the violin sound compared to the speech sound ($F$ (1, 118) =17.54, $p$ < .01). The advanced pronunciation group demonstrated a right-hemispherically dominant MMN to music durational stimuli ($F$ (2, 236) = 5.37, $p$ < .001), while the less-advanced group did not demonstrate any lateralization in their MMN strength. These results may explain that musical stimuli could have a predominant role in pre-attentive duration processing, and that the advanced pronunciation group and the less advanced group process music and language durational stimulus by activating different parts of the brain and with different degrees of intensity.

In addition to the results and implications of the duration discrimination task, a correlation was found between general musical aptitude (the total Seashore test score), all the Seashore subtest scores (pitch, rhythm, timbre, and tonal memory accuracy), and English pronunciation skill (-.54 < $r$ < .73, all $p$ < .001). Moreover, their study demonstrated that the general musical aptitude and pronunciation skills correlated with the performance IQ test ($r$ = .46, $p$ < .001), and the general intelligence test ($r$ = .32, $p$ < .005). This study implies that musical aptitude, English pronunciation skill, and general IQ are interconnected, and that the musical features of the linguistic stimuli could have an important role in the pre-attentive duration process.

## 2.4 Relationship between Language Prosody Processing Skill and English Proficiency

This section introduces the literature relevant to the correlation between language prosody processing skill and English proficiency, particularly with English listening proficiency.

**2.4.1 Linguistic sound sensitivity and EFL proficiency.** The Auditory perceptual acuity tests in the Hi-LAB (Doughty et al., 2010) and PLAB (Pimsleur, Reed, & Stansfield, 2004), which were developed for the theoretical background that phonetic discrimination and categorization are correlated with English proficiency, will here be introduced. Among various cognitive tests in the Hi-LAB, the Phonetic discrimination, i.e., Hindi-English Pseudo-Contrastive test, and the Phonetic categorization test, i.e., Russian test, were mainly introduced and discussed in Linck et al. (2013). They examined potential cognitive predictors of successful learning at advanced proficiency levels using this original language aptitude measure. The purpose of Linck et al.’s (2013) study was to obtain empirical evidence for the credibility of a language aptitude test, i.e., Hi-LAB.

Hi-LAB does not measure phonetic coding ability, which is important for reading, nor grammatical sensitivity, since these measures were considered relevant to an initial stage of second language acquisition (SLA). Hi-LAB focuses on cognitive and perceptual abilities, which may enhance language input processing and facilitate mapping in the memory of forms, meanings, and functions. Auditory perceptual acuity is the capacity to pay attention to and discriminate among speech cues. In this battery, the native-language processing system may not help detect the difference, so that L1 influence should be overcome, in order to hear, distinguish, and produce L2 speech.

The Phonemic discrimination task in Hi-LAB tests the ability to resist the tendency to assimilate new language sounds into already-possessed L1 categories. One task uses foreign contrastive phonemes that are also contrastive in English, such as the voiced /dz/ and an unaspirated /tʃ/ phoneme. The other difficult example task uses Hindi phonemes that fall into a single phoneme category in English. For example, the Hindi-English pseudo-contrastive test measures the ability to hear the difference between Hindi voiced /ǰ/ and voiceless /č/. These sounds correspond with the initial sounds of “jeep” or “cheap,” respectively. However, the stimuli have VOT values ranging from 120 ms to 0 ms, which are normally within the range of the English /ǰ/ phoneme (“j” sound). English speakers are expected to have difficulty distinguishing between these sounds.

These tasks are designed for L1 English speakers aiming at FLA for joining the military. The subjects were in various U.S. government agencies or members of the U.S. military. For this military recruitment purpose, high-level attainment was defined as having a highly proficient performance on the Defense Language Proficiency Test (DLPT; Defense Language Institute Foreign Language Center, 2009) and/or high-level job performance in a single language or multiple languages. The DLPT is a standardized proficiency test with sections and scores that assess listening and reading comprehension of a foreign language.

Linck et al. (2013) demonstrated that high-level attainment in listening or reading in DLPT was related to better Hi-LAB scores. Across all tested skills, Hi-LAB classification accuracy between high-attainment and mixed attainment groups ranged from 58.8% to 72.1% (Linck et al., 2013, p. 555). The other result from the validity study (Doughty, 2013, p. 163) indicates that Hi-LAB correctly classified high-attainment learners with up to 70% accuracy when examining listening proficiency, and nearly 60% for reading proficiency.

Compared to Hi-LAB, PLAB has four common constructs: Phonetic code ability, grammatical sensitivity, rote learning ability, and inductive language learning ability. The PLAB also includes a measure of phonemic discrimination. This battery constitutes six categories accounting for 117 points in total, two of which are tests related to auditory ability. One is a 30-point test for sound discrimination, i.e., the ability to learn new phonetic distinctions and to recognize them in different contexts. The other is a 24-point test for sound-symbol association, i.e., the association of sounds with their written symbols. Pimsleur (1966) structured PLAB with the four factors significantly related to foreign language learning: grade point average, motivation, verbal ability, and audio ability. The audio ability score accounts for nearly half of the total score of the battery (54 out of 117 points in total). Hi-LAB and PLAB measure audio perceptual acuity at the phonological level of words. The reliability studies (including a validity component) of Hi-LAB as a measurement of English proficiency is highly reliable (alpha = .90) (Doughty, 2013). This high reliability of Hi-LAB and PLAB in measuring English proficiency and audio perceptual acuity implies that audio perceptual acuity likely contributes to English proficiency.

**2.4.2 Tonal/phonological short-term memory and English listening comprehension skill.** Call (1985) studied the relationship between tonal/phonological short-term memory (STM) and EFL listening comprehension skill, in terms of the input hypothesis.[12] The participants were 41 students enrolled in the English Language Institute of the University of Pittsburg whose native language was Arabic or Spanish. They cooperated in taking the Michigan Test of Aural Comprehension (Upshur et al., 1969) as a measure of processing ability of English auditory input, and the tonal memory section of the Seashore Measures of Musical Talent (Seashore et al., 1960). Auditory memory was measured by (1) sentences taken from running discourse, i.e., probe, (2) individual sentences, (3) random words, and (4) random digits.

The correlation analysis revealed a statistically significant correlation between listening comprehension skill and (1) sentence STM ($r$ = .65), (2) probe STM ($r$ = .57), and (3) tonal memory ($r$ = .42). The partial correlation analysis demonstrated that variables such as sentence memory, story memory, and tonal memory accounted for 47% of listening comprehension ($R^2$ = .47). Out of this 47%, sentence memory alone accounted for 42% of listening comprehension ($\Delta R^2$ = .42), probe memory for 1% ($\Delta R^2$ = .01), and tonal memory for 4% ($\Delta R^2$ = .04).

The study suggested that memory for syntax, which was measured by the sentence or probe subtests, can predict English listening comprehension skill. In contrast, the variables of memory for random words and digits did not contribute to explaining listening comprehension, probably due to their incomprehensibility. Call (1985) implied the importance of teaching syntactic structure for auditory recognition, for such a structure gives shape to the meaning. Moreover, it can be inferred from the study that enhancing tonal memory through music training might affect listening comprehension in English.

**2.4.3 Correlation between accent sensitivity and English listening comprehension.** Meerman et al. (2014) investigated the degree to which accent sensitivity provides the foundation for lexical knowledge and listening comprehension for Japanese university students learning EFL. Accent sensitivity was measured with 40 sets of correctly and incorrectly pronounced words, half nouns, and half adverbs, in a total of 80 words taken from a TOEIC preparatory textbook. Listening comprehension was tested with multiple-choice questions on six short dialogues from TOEIC preparatory materials. Lexical knowledge was tested with 40 multiple-choice questions including three incorrect explanations and one correct explanation. Unexpectedly, the authors concluded from the structural equation modeling analysis that accent sensitivity makes little contribution to either participants' lexical knowledge or listening comprehension. They induced from this result that Japanese English education does not pay sufficient attention to teaching accent sensitivity to further contribute to listening comprehension. Future study should be conducted after an appropriate and adequate accent sensitivity training is provided to see how acute accent sensitivity correlates with English listening comprehension.

### 2.5 Music and Language Specified Areas in the Brain

Some researchers have suggested that music and language have similarly structured syntaxes and share the cortical network that is involved in the learning of speech and tone sequences (Patel & Daniele, 2003; Pesetsky, 2008; Maess et al., 2001). ERP studies have been conducted in the field of neuroscience, psychology, education, and linguistics, in order to reveal how the cortical network functions. A recent ERP study (Koelsch et al., 2002) found that “brain responses reflecting the processing of musical chord-sequences were similar to brain activity elicited during the perception of language, in both musicians, and non-musicians.”

Koelsch et al. (2002) revealed, using functional magnetic resonance imaging (fMRI) data, that the human brain employs the certain neuronal network not only for language processing but also for processing musical information. This may suggest that the cortical network known to support language processing may be a shared domain for music processing.

The neuropsychological evidence indicates that some brain regions (the inferior frontal cortex including Broca’s area) often assumed to be language-specific are implicated in musical processing (Levintin & Menon, 2003). Maess et al. (2001) used magnetoencephalography (MEG) and found that Broca’s area and its right-hemisphere homologue are involved not only in syntactic analysis during auditory language comprehension but also in the analysis of incoming harmonic sequences, i.e., musical syntax. Specifically, harmonically inappropriate chords activated Broca’s area and its right-hemisphere homologue. This finding demonstrates that “complex rule-based information is processed in these areas with considerably less domain-specificity than previously believed, which might suggest that these areas process syntax, that is, complex rule-based information in non-linguistic domains such as musical or other auditory domains” (p. 543). This observed finding secondarily implies the certain relationship between the processing of complex rule-based language and such a music, which may partly account for the influence of music training on the processing of unfamiliar foreign languages by activating the shared domain in a brain.

With regard to brain localization, Jackendoff and Lerdahl (1983) remarked that musical sophistication is, in fact, associated with an advantage in both hemispheres, whereas it is widely thought that the left hemisphere of right-handed people subserves linguistic functions and the right hemisphere subserves musical perception. ERP studies have also been conducted on linguistic prosody. For example, Steinhauer et al. (1999) applied ERP measures to reveal the time course and neural basis of prosodic information processing. They reported that electrical brain responses to unexpected items in a structured sequence, i.e., hierarchical structure, were often lateralized to the left in processing language. Friedrich (2004) additionally reported that the brain responses were often lateralized to the right when processing prosody, such as rhythm, pauses, accents, amplitude, and pitch variations. If prosodic and music pitch perception rely on a common neural system, durational violations may elicit similar effects for both language prosody and music. However, an activation of the shared brain region for prosody processing and music processing may depend on the types of structural information (Friedrich, 2004).

Some of these studies imply that the brain region that reacts in both language and music prosody processing may be activated by auditory training. Some researchers have discussed whether musical ability could be a predictor of high SLA competence (Gottfried, 2007). Since the correlation between the prosody processing of language and music seems to exist based on the findings from the ERP studies, the hypothesis can be constructed that linguistic sound sensitivity is correlated to musical sound sensitivity. This hypothesis will be strengthened by the research on the musical and linguistic prosodies of Specially Language Impaired (SLI) children in the next section.

## 2.6 Functions of Language Prosody Revealed in Studies with SLI and Autistic Children

Based on the deductions from the studies reviewed above that there exists a correlation between the processing of language and music, this section explores such a correlation as uniquely observed in the language and music of SLI and autistic children. These groups are reported to have deficient rhythmic prosodic ability in their language use (Weinert, 1992). Their impairment implies that rhythmic prosodic ability is required for achieving intra/interpersonal communication. Children with autism are also characterized as having low musicality and low prosody processing skills (Peppé et al., 2007). Peppé et al. (2007) described both receptive and expressive prosodic impairments in children with autism, and demonstrated the clear relationship between language prosody processing skills and language proficiency.

As already described in Chapter 2, children with Asperger's-type autism have symptoms that may include high verbal ability, low sociability, and low musicality (Huron, 2003). Comparatively, children with SLI have low verbal ability, low sociability, and low musicality (Weinert, 1992) and cannot fully command their native languages. Weinert (1992) conducted an empirical study on English speaking SLI children, and concluded that they lacked rhythmic perception of English sentences. The experimental procedure was as follows:

In the experimental task and design, the participants were asked to learn 12 structured strings of a miniature language[13] and were tested for its rule knowledge where they had to recall a new set of grammatically well-formed and rule-violating strings. In the reproduction tasks of the miniature language, children with SLI tended to have more difficulties in learning prosodically enriched phrases than monotone phrases. Compared to SLI children, the children who were able to learn more prosodically enriched sentences in the miniature language were also able to judge the grammaticality of new strings, and demonstrated better performance on the rhythm discrimination tasks[14] (Weinert, 1992).

Weinert (1992) suggested that input language must include a variety of cues as to its structural regularities, and that the learner must have the ability to use these cues in rule induction in order to acquire a complex rule system successfully (p. 568). Through his experiment, Weinert (1992) reported that SLI children have deficits in the ability to process and use the rhythmic-prosodic structure of speech as cues to formal grammatical regularities. In other words, these deficits impede the implicit learning of the formal regularities of an input language.

Weinert (1992) added that prosodic cues are important, since the processing of prosodic cues in an input language does not depend on preliminary syntactic or semantic analysis but it may be the basis of such analysis. That is, learners who are not yet knowledgeable about the semantic or syntactic rules of a target language can acquire clues for comprehension from these prosodic cues. Some children who have difficulties in a foreign-language learning environment might not yet have become sufficiently sensitive to its prosodic cues. Deficits in prosody processing ability may lead to defective language acquisition.

Language prosodic cues are specifically considered as pausing (Goldman-Eisler, 1972), segment lengthening (Cooper & Paccia-Cooper, 1980; Klatt, 1975), pitch discontinuities (Cooper & Sorenson, 1977), and the syncopation of rhythm (Scott, 1982). The grammatical functions of prosody include the segmenting of utterances into prosodic phrases. The ends of phrases are signaled (1) by a number of prosodic factors, including a pause after the phrase (Butcher, 1981);

(2) by lengthening of the final syllables (Scott, 1982); and (3) by inclusion of a nuclear tone located at or near the end of the utterance (Crystal, 1969). These prosodic cues listed above seem crucial for communication; they are treated as prosodic variables measured in Studies 1-3 in the present thesis.

After reviewing the relevant previous research in Chapter 2, a proposal can be made regarding the relationships between (1) sound sensitivity for perception and production of music and language; (2) sound sensitivity and L1 prosody processing ability; (3) sound sensitivity and L2 prosody processing ability; and (4) English prosody processing ability and English proficiency. Supported with a brain study and an SLI study, sound sensitivity and English prosody processing skills appear to have the potential to affect each other in order to realize high English proficiency.

In the next section, previous academic studies on the sensitive period (Long, 2013) in which sound sensitivity and English prosody processing skills can be acquired through training or learning will be reviewed and discussed, in order to achieve native-like English proficiency as an ultimate goal of EFL.

## 2.7 Supposition of Early Effect on Prosody Acquisition in EFL Programs

The studies discussed above suggest that there is a correlation between sound sensitivity, language prosody processing ability, and language proficiency. However, there is no research that the author knows of that actually ties all these three together well, and thus, this is the main job of this dissertation. In this section, previous studies on the sensitive period for the acquisition of sound sensitivity for processing language and music will be reviewed.

With regard to ultimate language acquisition, it has been proposed that there may be a sensitive period in which learners can acquire sound sensitivity for music and language. For example, a number of early sensitive hypotheses and critical period hypotheses for language acquisition have been formulated (Krashen et al., 1979; Slavoff & Johnson, 1995; Long, 1990a). Chang and Merzenich (2003) reported from their empirical studies that the normal development of the auditory cortex of rats within a two–three week critical period was biased by the specific environmental inputs. They deduced that human infant L1 acquisition is “a manifestation of this powerful sound-exposure-based critical plasticity” (p. 499). Their study implies that successful language acquisition is realized through this adaptability, but unsuccessful acquisition can also be realized through the impediment of expectedly quantified sound exposure.

According to Long (1990a), when learners are at a “critical” age for second language acquisition, they will be able to develop native-speaker like ability: This periodical limitation for

FLA is hypothesized as the critical period hypothesis (Long, 1990a). Long's (1990a) description of the critical period is as follows: The ability to attain native-like phonological abilities in FLA (1) begins to decline by age six in many individuals, and (2) is impossible to acquire beginning later than age 12, no matter how motivated a person might be or how much opportunity they might have.

In the critical period for FLA, children have an advantage not in rate of learning but in ultimate attainment (Slavoff & Johnson, 1995), and the decline in foreign language learning ability takes place gradually before puberty, at age 6–7, and drastically around puberty, at age 15–17 at the latest (Johnson & Newport, 1989). DeKeyser (2000) provided evidence for Bley-Vroman's (1988) *Fundamental Difference Hypothesis*[15] by demonstrating that no adults reached a native level of competence in L2 morphosyntax without relying on explicit, analytic, problem-solving capacities. Moreover, his study demonstrated that age and aptitude interact in the sense that (a) age of start makes a clear difference for those who have average or below-average verbal ability and (b) verbal ability makes a difference for those who start to learn a foreign language as adults (DeKeyser, 2000, p. 518). DeKeyser (2000) described the "maturational effect" in comparison with the age effect. That is, after childhood, explicit learning processes are a necessary condition for achieving a high level of competence in a nonnative language.

In spite of the age-success correlation shown in the figure by Doughty et al. (2010, p. 11), some individuals were reported to attain high-level proficiency in a foreign language with a late start. Doughty et al. (2010) proposed that language aptitude could make up for the post-critical period disadvantages in FLA.

According to the studies discussed above in this section, it seems preferable for learners to start no later than six years old to obtain native-like fluency in a target language. However, even within this early period, the input volume should equal that provided by an immersion program. Dominguez and Pessoa (2005) presented empirical findings that support the effectiveness of early-started foreign language learning from pre-school age. They made an empirical comparison between early learners that started learning Spanish in kindergarten and learners that started in their sixth year. Their findings were based on the results from a battery of oral and written tests on Spanish proficiency, where the early-start learners outperformed the late-start learners. If young learners have a higher potential to acquire native-like pronunciation of a target language than adult learners do, as the above research suggests, then those who began being exposed to the language earlier within the critical period will have a greater advantage. Nevertheless, the volume of target language input will have an impact. The process requires massive amounts of input, which only a total immersion program can provide, not a program consisting of a few hours of foreign language teaching per week.

Yusawa, Sekiguchi, and Rie (2007) reviewed the previous research on young children's acquisition of EFL. Their research suggests that (1) younger children are better able to learn the phonology of a foreign language; (2) phonological categories of the L1 remain flexible in early childhood; (3) young Japanese children are able to perceive English sounds in a consonant-vowel, i.e., a CV structure, correctly; and (4) young Japanese children are sensitive exclusively to the boundaries between mora, which makes it difficult to perceive English sounds in a VC structure or in multiple syllables (p. 153).

Yusawa et al. (2007) state that children's facility in English phonological awareness is closely related to their acquisition of English vocabulary. Muñoz (2010) suggested that the preferable conditions for early learning were that (1) the amount of exposure to the target language should be determined; (2) learners should participate in a variety of social contexts consisting of different age groups; and (3) educators should be informed of what is expected to be provided at each step of learning for each age group.

In Japan, public English education is supposed to begin from 11 or 12 years old, although this will be reformed by 2020 as 9–10 year old students having English language activity classes

1–2 times a week (MEXT, 2015a). When children start to learn foreign languages from their fifth year of elementary school, it is not very easy to acquire the language implicitly, if relying theoretically on the critical period hypothesis. Even a Japanese child with good English reading skills cannot always communicate spontaneously in daily conversation. An explicit learning mechanism may make it challenging for older learners to gain native-like command of foreign languages.

Krashen et al. (1979) supported the *early sensitivity hypothesis* (p. 573): Acquirers who begin to be naturally exposed to an L2 in childhood tend to achieve higher L2 proficiency than those who begin learning as adults. Krashen et al. (1979) made two generalizations on the SLA process:

1. Adults proceed through early stages of morphological and syntactical development faster than children do (when time or exposure are held constant).

2. Older children acquire an L2 faster than younger children (again in the early stages of morphology and syntax, where time is held constant).

However, they added that “adults and older children in general initially acquire the L2 faster than younger children, but a child L2 acquirer will usually be superior in terms of ultimate attainment” (p. 574). Turnbull, Lapkin, Hart, and Swain (1998) and Dominguez and Pessoa (2005) also indicated, with regard to the advantages held by children in FLA, that beginning foreign language instruction at an early age may be advantageous for developing students' oral skills and their confidence in using the target language. Early foreign language learners may have a cognitive, academic, and attitudinal advantage in oral communication, compared to older learners.

These early sensitivity hypotheses imply one of the advantages for early FLA starters, which is that they may have acute auditory sensitivity in processing a foreign language. In language aptitude tests (Pimsleur, 1966; Doughty et al., 2010), audio processing of phonemes is regarded as a prerequisite constituent of linguistic aptitude in FLA. Thus, if late FLA starters try to reach the ultimate level of a target language, they must have obtained this aptitude as a precondition. Various periodical factors in FLA, such as a critical period or a sensitive period, and the age of onset, affect pronunciation of a target language to a great extent (Long, 1990b). English sound sensitivity, i.e., English prosody processing skills, seem to be a crucial factor for native-like pronunciation and comprehension. In the next section, preferable ages for music sound sensitivity will be overviewed, with reference to previous studies.

## 2.8 Contribution of Early Music Education to Enhancing Sound Sensitivity

Auditory perceptual acuity tested in phoneme discrimination and phonemic categorization (Doughty et al., 2010), and in phonetic distinction and sound symbol-association (Pimsleur, 1966) is considered a significant factor in language aptitude tests. Pitch perception (Marques et al., 2007; Schön et al., 2004) and rhythm perception (Nakano & Natsume, 2011) are also considered to be promoting factors in language acquisition. As the equivalent of such an aptitude test in music, which tests musical tone and rhythm sensitivity, Musical aptitude tests have been developed by Seashore et al. (1960), Wing (1968), and Gordon (1982).

The early training effects and a few critical period theories of music will be introduced in this section. Less research, however, has been conducted on the critical period for musical acquisition, except with regard to *absolute pitch* (Crozier, 1997; Russo, Windell, Weinert, & Cuddy, 2003). Russo et al. (2003) suggested that the critical period for the acquisition of absolute pitch is age five–six years, based on their experimental study. With regard to the music sensitive period, Gordon (1982) states that the potential to achieve musical sound sensitivity stops developing after the age of approximately nine years, and then stabilizes and remains consistent throughout life. Thus, Gordon suggested the importance of the quality of classroom and private music instruction, particularly during the first three years of elementary school.

Shahin, Roberts, and Trainor (2004) followed the musical development of children from the beginning of their musical lessons at age four–five and compared their development to age-matched non-musician children. They found larger P1, N1, and P2 components[16] in the auditory evoked potential (AEP) response to violin, piano, and pure tones both just prior to the music lessons and one year later, in the musician children than the non-musician children (aged-matched controls). The AEPs observed for the instrument of practice in musicians were nearly equal to those of non-musician children approximately three years older. This AEP study suggests that the effectiveness of music lessons for tone sensitivity can be observed at age four–five years.

Up to this penultimate section of Chapter 2, previous studies on the sensitive period for language/music sound have been described. The remarkable implications on the influence of musical sound sensitivity to language prosody processing can be derived from the study results that humans have greater sensitivity in music to tonal incongruities than sensitivity in spoken language to pitch incongruities (Schön et al., 2004), and that musical stimuli could be more predominant in pre-attentive duration processing than language stimuli (Milovanov et al., 2009). The following section presents the research questions of this study in light of the previous research described above.

**2.9 Proposals and General Purposes of Experimental Studies 1–4**

Summarizing the previous research in Chapter 2, initially, the concepts and functions of music and language were explained, specifically with regard to their common points and correlations. Secondly, previous research on (1) the relationship between sound sensitivity and L1/L2 language prosody processing skills, (2) the relationship between sound sensitivity and phonological awareness, and (3) the relationship between English prosody processing and English proficiency in EFL were reviewed. Thirdly, research on (1) the specified brain areas for language and music processing, (2) the functions of language prosody as revealed in studies with Specially Language Impaired (SLI) and autistic children, and (3) the effectiveness of early prosody acquisition in EFL programs and early music education were examined.

Through investigation of previous research in Chapter 2, it has become apparent that audio processing of phonemes is regarded as a prerequisite constituent of linguistic aptitude in FLA (Pimsleur, 1966; Doughty et al., 2010). The study of phoneme recognition is categorized as segmental phonology, while that of prosody processing is categorized as supra-segmental phonology (Markus, 2005). While English phoneme processing has been recognized as a part of English aptitude, general sound sensitivity and the auditory processing of English prosody as contributors to English aptitude have been under-explored. Thus, the objectives of this dissertation are to investigate (1) if sound sensitivity can affect the processing skill of English prosody, and (2) if English prosody processing skills can affect English listening comprehension. Moreover, this study seeks to clarify (3) if elementary and high school participants are significantly different in their sound sensitivity and English prosody processing skills, and (4) if Western instrumental music training improves musical sound sensitivity. With regard to the third point, the elementary school participants have been taking school English classes for approximately three years from the first grade rather implicitly, while high school participants have been taking school English classes for three–five years from junior high school rather explicitly. Thus, comparison between these two groups will imply how different the outcome can be between early and late English learning and between implicit and explicit English learning.

The specific implications have been provided by previous research that: (1) sound sensitivity may be one kind of language aptitude; (2) English prosody processing skills may correlate with English listening proficiency; and (3) sound sensitivity may contribute to English listening comprehension and production implicitly. Unfortunately, few research studies by Japanese EFL faculties have examined the effectiveness of English prosody education in Japan, nor the correlations between English prosody processing skill and musical sound sensitivity of

Japanese EFL learners. The early acquisition of English prosody perception/production and sound sensitivity has not been discussed sufficiently in the Japanese context. Considering this reality, this thesis can find some clues to improve Japanese school English education and music education by answering the research questions in 2.10 that follow.

## 2.10 Research Questions

The ultimate objective of this study is to clarify the hypothesis that sound sensitivity, i.e., music audiation, in particular tonal/rhythmic perception and rhythmic production, is related to receptive/productive prosody processing skill of English, and consequently English listening comprehension skill. If this first hypothesis is asserted, it will lead to the deduction that learners' sound sensitivity implicitly affects their English listening comprehension skill. The second objective is to ascertain if musically trained children tend to have higher sound sensitivity than untrained children do. Musically trained children are defined as children with three years of music training for more than 30 minutes per day, before their 10th birthday.

The research questions have not yet been clarified in previous research, as shown through analysis of the relevant previous studies in Chapter 2. In order to achieve the objectives of this thesis, the following research questions will be answered in each experiment in the following chapters:

1. Does sound sensitivity contribute to prosody perception/production of English as a foreign language?

2. Does sound sensitivity implicitly contribute to listening comprehension in English as a foreign language?

3. Does English prosody perception/production contribute to English listening comprehension in the process of foreign language acquisition?

4. Do early musical lessons (between the ages of three and nine, with daily music training for more than 30 minutes) affect sound sensitivity?

In Chapter 3, the correlations between sound sensitivity, English prosody processing skill, and English listening proficiency will be examined through four experimental studies. The first to third experimental studies will be conducted with students in the 8–10 age group and the 15–17 age group, in order to compare those three variables between elementary learners who have just started receiving compulsory education and senior learners who have mostly completed it.

The fourth experimental study will be conducted with 6-12 year old participants who have different levels of musical experience, in order to compare the effect of music training on their sound sensitivity.

# Chapter 3 Study 1: Relationship between Sound Sensitivity, English Prosody Perception, and EFL Listening Comprehension of High School Students

## 3.1 Purpose of Study 1

Experimental Study 1 was conducted to clarify research questions 1–3: (1) With regard to high school students, does sound sensitivity contribute to prosody perception of EFL?; (2) Does sound sensitivity contribute to English listening comprehension?; and (3) Does English prosody perception contribute to English listening comprehension in EFL?

## 3.2 Participants

The participants comprised 75 students at a Japanese public high school, 20 of whom were students in their 10$^{th}$ year and 55 of whom were students in their 11$^{th}$ year. The participants comprised 16 males and 59 females attending a public vocational high school, where students major in either agriculture or domestic science.

## 3.3 Materials

The following three materials were used for experimental Study 1 with 75 high school students.

1. Measurement of language prosody perception: PEPS-C (Peppé & McCann, 2003)

This test measures receptive skills in English prosody and has three subtests that measure specific skills such as intonation, chunking, and focusing. Each test takes approximately 10 minutes to complete. This prosody measure is adaptable to those who have basic English skill at a level above four-year-old native speakers of English. The PEPS-C contents are as follows:

a. Prosody input test

This subtest assesses the ability to perceive prosodic differences. Participants identify whether the prosody of pairs of phrases is the same or different. The prosodies were low pass filtered sound signals of English phrases, which sounded like noises consisting of more than two words. This test includes 16 questions. In the sample low pass filtered phrases in Figures 1 and 2, the prosodic boundaries can be marked with a slash (a small boundary) or double slashes (a big boundary) in the original phrases as follows: Fruit / / salad / / and milk (Figure 1); fruit / salad / / and milk (Figure 2).

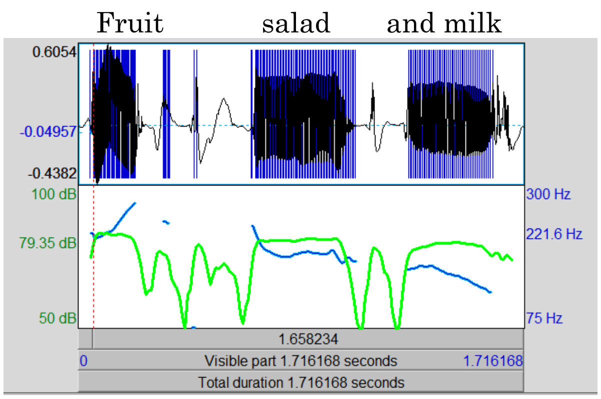



Figure 1 “Fruit, salad, and milk”

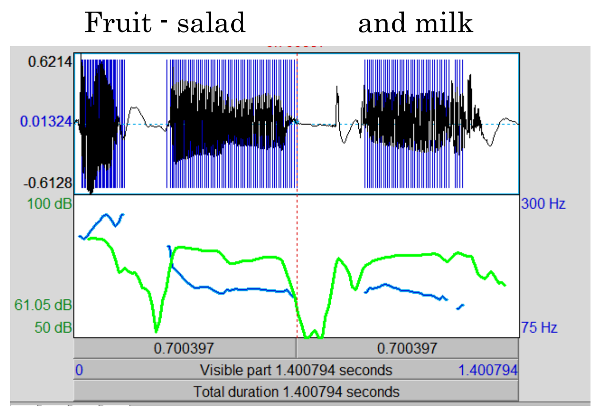



Figure 2 “Fruit-salad, and milk”

*Figures 1 and 2.* Low pass filtered English phrases with a cut-off frequency of 300 hertz, using Praat software.

b. Chunking input test

This subtest assesses the comprehension of prosodic boundaries of phrases, such as “pink-and-black and green socks,” “pink and black-and-green socks,” and “chicken fingers and fruits.” In this subtest, participants are shown a pair of pictures and they listen to a phrase with prosodic boundaries. Participants are then asked to choose the picture that illustrates the pronounced phrase. This test includes 18 questions. Figures 3 and 4 show the intensity, the pitch, and the pauses between blue colored sound waves. The prosodic boundaries can be marked with a slash (a small boundary) or double slashes (a big boundary) as follows: Chocolate / ice / cream / / and honey (Figure 3); chocolate / / ice-cream / / and honey (Figure 4). See Appendix A1 for the test script.

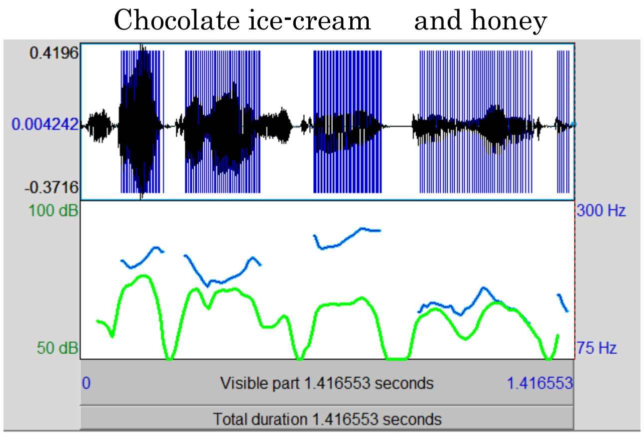



*Figure 3.* “chocolate ice-cream, and honey.”

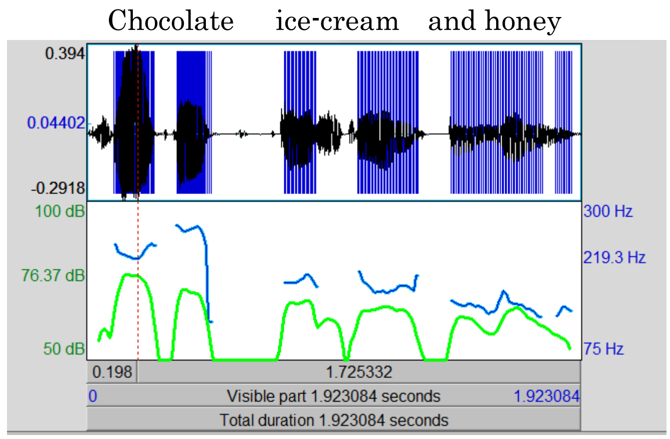



*Figure 4.* “chocolate, ice-cream, and honey.”

c. Focus input test

This test assesses the ability to identify contrastive stress (focus). In this subtest, participants are shown a pair of pictures and listen to a sentence with a focus on a particular word. Participants are then asked to choose the picture that illustrates the focus word. For example, participants hear the sentence “I wanted blue and BLACK socks (capital letters indicate stress)” while looking at pictures of black socks and blue socks. The sentence means, “I forgot to buy the black socks,” so the participants should choose the picture of black socks. This test includes 16 test questions. See the Praat figures in Figures 5 and 6 as well as Appendix A2 for the test script.

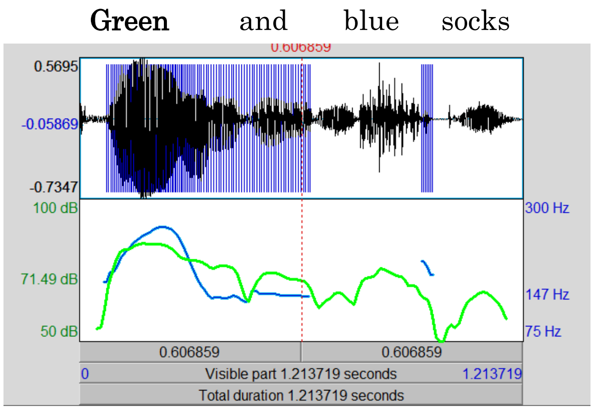



*Figure 5.* “**green** and blue socks.”

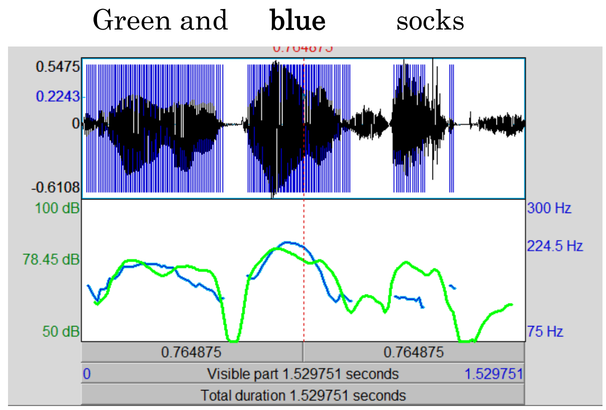



*Figure 6.* “green and **blue** socks.”

2. Measurement of specific language prosody: Prosody discrimination test (original test created by the author)

This test was created specifically for the purpose of this study, to measure the ability to discriminate between the prosody of American/Australian English and nine other languages including Chinese, Russian, French, Arabic, Polish, Korean, German, Mongolian, and Japanese. Rhythmically classifying these languages using the nPVI (Grabe, 2002), Arabic, American/Australian English, German, and Russian are classified as stress-timed languages, in which the temporal duration between two stressed syllables is mostly equal. Polish is classified as either stress-timed or syllable-timed. French, Mandarin Chinese, Mongolian, Korean, and Japanese are classified as syllable-timed languages, in which the duration of every syllable is traditionally considered to be approximately equal. Japanese is more precisely classified as mora-timed language, in which the duration of every mora is considered equal.

*Test development procedure.* Firstly, 12 English sentences were recorded by five native speakers and the author, including two male Americans, two female Americans, one male Australian, and a Japanese female with a TOEIC score of 860. The sentences were composed by the author, based on an average sentence length of nine words (6–14 words) and including a variety of sentence forms such as declarative, directive, and interrogative. The sentences were checked by native speakers. Twelve English sentences were translated into one of the nine other languages by native speakers of those languages. The translated sentences were then recorded by the native speakers of each language. The 12 pairs of sentences were pronounced by the same gender since fundamental frequency is similar between voices of the same gender.

All the recorded sentences were low pass filtered with a cut-off frequency of 300 hertz, created by the software Praat. In such pairs of sounds, participants could not hear the words but only the prosody. Figures 7 and 8 demonstrate the different tone and rhythm structures of Mandarin Chinese and English.

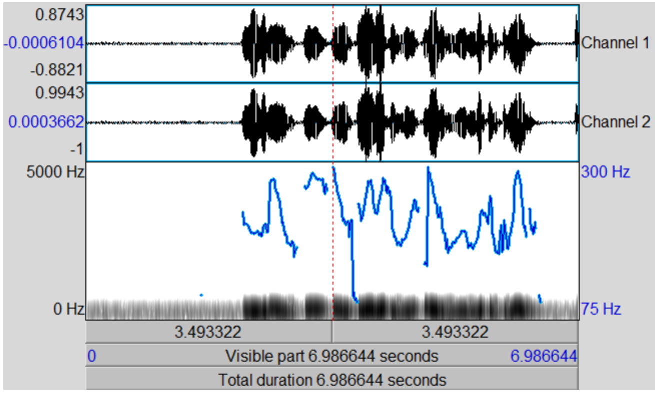



*Figure 7*. “我们的老师在情人节送给了我们巧克力”: In Pinyin; wǒ mén de lǎo shī zaì qíng rén jíe sòng4 gěi le wǒ mén qǐao kē lì. See the Glossary for an explanation of Mandarin Chinese tones and diacritics.

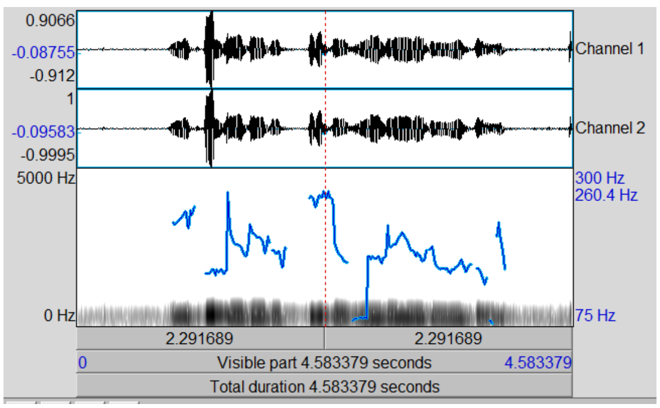



*Figure 8.* “Our teacher gave us chocolate on Valentine’s Day.”

The figures above show low pass filtered Mandarin Chinese (Figure 7) and English sounds (Figure 8) with a cut-off frequency of 300 hertz, created with Praat software. As shown in Figures 7 and 8, in English, the first syllables of “our,” “teacher,” “gave,” “chocolate,” “Valentine’s,” and “day” have accented peaks, and create a contour, compared to Mandarin Chinese, in which all 17 syllables retain almost the same pitch value. In Mandarin Chinese, a constant relationship exists between the shape and range of the six contrastive tones (Fok, 1974).

*Test application procedure.* Participants listened to 12 pairs of low-pass filtered sentences and identified which one of the pair resembled English. The voice gender of sentences was the same: There were seven male voice sentence pairs and five female voice sentence pairs. Appendix A3 provides details of the 12 sentence pairs used in this test, which takes approximately eight minutes to complete. Appendix A4 provides the answer sheet for the participants to fill in. Appendix E provides the Praat graphs of the 12 sentences.

3. Measurement of listening comprehension: Eiken Grade 3 listening test

The Eiken Grade 3 listening test is aimed at Japanese junior high school graduates. Examinees are expected to be able to understand and use language concerning familiar, everyday topics, such as likes and dislikes, and basic personal and family information. The examination is divided into two parts: Stage 1 (a written examination including a listening section) and Stage 2 (an interview-format speaking test). On this occasion, only the listening test was administered. This test consists of (1) 10 choice questions on short conversations taking place between two persons in a picture, with three answer choices for each question, (2) 10 choice questions on four sentence conversations between two people, with four answer choices for each question, (3) 10 choice questions on the passage contents, with four answer choices for each question. It takes approximately 30 minutes to answer 30 questions. Example questions are shown in Appendix B.

4. Measurement of Musical Audiation: “Intermediate Measures of Music Audiation” (Gordon, 1982)

This test is a measure of the musical tone and rhythm processing skills of schoolchildren. It consists of 40 tone samples and 40 rhythm samples. Each tone sample in a pair consists of three tone series, which may be the same or different. The test questions are presented with the synthesizer. The participants must identify whether the pairs of tones are the same or different. Each rhythm sample in a pair consists of notes of one measure, which may be the same or different. The participants must identify whether the pairs of rhythms are the same or different. Each takes 10 minutes to complete.

The Tonal test measures the audiation (hearing music in mind) of keyality, i.e., tone (C, D, E, F, G, A, H); and the Rhythm test measures the audiation of tempo, i.e., the relative length of beats in terms of speed. Only these two basic dimensions of developmental musical aptitude are tested, which is sufficiently short and simple to be validly administered to young children.

To be precise, in the Tonal test, each of two phrases in a question, composed of one tonal pattern of three tones in length, is performed at the same tempo and in the same keyality. That is, in the Tonal test, the tones in the phrases are performed in beats of equal length. At least one phrase of the pair in each tonal question includes the tonic; for example, D is the tonic in the D keyality, while C# is the tonic in the C# keyality.

In the Rhythm test, each of two phrases in a question is composed of one rhythm pattern: Notes in the phrases in the Rhythm test are performed at the same pitch. The majority of questions on the Rhythm test of the *Intermediate Measures of Music Audiation*, which are composed of different phrases, are in the same tone.

Theoretically, this audiation depends upon children's informal musical achievement based on musical aptitude. See Appendix C for Tonal test notes and Appendix D for Rhythm test notes.

5. Survey of music learning background and English learning background: Questionnaire

The questionnaire contains questions on years of music training, instruments or areas of training, starting year, training hours, motivation to train, and if in possession of absolute pitch. It also contains questions on starting year and period of EFL training at home and at school. Other questions concerned the studying hours, motivation, interest in EFL, and degree of perception and production of English. See Appendix F for the script.

## 3.4 Procedures

The research procedure is described in steps 1 to 5 below.

1. In the fourth week of March 2014, the two major elements of musical skills, which are tonal and rhythmic processing skills, were measured by administering the Intermediate Measures of Music Audiation test (Gordon, 1982) in participants' homeroom classes in one day for each class.

2. In the fourth week of March 2014, the participants' English prosody processing skill was measured with the Profiling Elements of Prosodic Systems–Children (Peppé & McCann, 2003), and English prosody discrimination skill was measured with the Prosody discrimination test, in their homeroom classes.

3. In the fourth week of March 2014, a questionnaire was administered on music learning background and English learning background.

In the third week of July 2014, the Eiken Grade 3 listening test was administered to the students in their homeroom classes.

## 3.5 Results

The statistical data will be shown and explanations will be given on: (1) descriptive statistics, (2) Pearson's correlation, (3) collective regression analysis/stepwise regression analysis, and (4) structural equation modeling (SEM). Before presenting the results of these statistics, uncontrolled factors such as age and gender difference in the same groups of participants were checked with independent *t*-tests. Following the analyses of musical sound sensitivity, a score difference due to gender was suspected. Thus, an independent *t*-test was performed on the tonal and rhythm scores of 75 high school students. No significant difference was found between the average tone scores of 16 males ($M$ = 27.4) and 59 females ($M$ = 33.8) ($t$ (73) = 1.95, $p$ = .07). Furthermore, no significant difference was found between the average rhythm scores of 16 males ($M$ = 29.0) and 59 females ($M$ = 32.5) ($t$ (73) = 1.22, $p$ > .24). As a whole, from the results of the *t*-tests, no significant difference in musical sound sensitivity was found between male and female high school students. The *t*-test tables can be seen in Appendices G and H.

Following the analyses of Eiken Grade 3 listening test scores, a score difference due to school age was suspected: The students in their first year of high school have a disadvantage in English listening comprehension, due to their having one year less immersion in English classes. Thus, an independent *t*-test was performed to compare listening comprehension between the students in their first year and those in their second year of high school. The result found no significant difference between the average Eiken Grade 3 listening test scores of students in the first year (*M* = 12.9) and second year (*M* = 11.8) ($t$ (73) = 0.96, $p$ = .34). The tables of *t*-test results can be seen in Appendix I. Since no significant difference was found in English listening comprehension between the first year and the second year students (in fact, the trend even belied the expectation; e.g., the older students performed worse than the younger ones), they can all be considered to belong to the same experimental group named high school participants.

### 3.5.1 Descriptive statistics.

Table 4

*Descriptive Statistics for the Tests Used in Study 1*

| Variable | *K* | *M* | *SD* | Range | *α* | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| Prosodic skill measurements: | | | | | | | |
| Focus input | 16 | 14.24 | 2.18 | 6-16 | .83 | -2.51 | 6.00 |
| Chunking input | 18 | 11.04 | 2.83 | 6-16 | .60 | -.99 | 2.40 |
| Prosody input | 16 | 14.72 | 3.19 | 2-16 | .91 | -2.71 | 7.07 |
| Prosody discrimination | 12 | 8.16 | 2.27 | 4-12 | .57 | -.82 | 1.58 |
| English listening skill measurement: | | | | | | | |
| Eiken Grade 3 listening test | 30 | 11.91 | 4.38 | 6-22 | .60 | -.023 | -.21 |
| Musical aptitude measurement: | | | | | | | |
| Tone perception | 40 | 32.47 | 8.54 | 0-40 | .96 | -2.53 | 6.40 |
| Rhythm perception | 40 | 31.75 | 7.50 | 0-38 | .86 | -2.68 | 8.07 |

*Note: K* indicates the number of items in each test.

Table 4 shows the descriptive statistics for all tests in this study, including mean, standard deviation, minimum score, maximum score, and test reliability (Cronbach's $\alpha$). The test reliability of the Chunking input and Prosody discrimination tests was rather low, with $\alpha$ values of approximately .60. The kurtosis of the Focus input test, the Prosody input test, the Tonal test, and the Rhythm test are rather high, compared to the Eiken Grade 3 listening test. These high kurtosis values for prosody skill measurements and musical aptitude measurements may mean that the distributions of those test scores are distorted toward relatively high scores.

**3.5.2 Correlation analysis between sound sensitivity variables, English prosody processing variables, and an English listening comprehension variable.** Table 5 shows the statistical correlation between musical sound sensitivity and English prosody processing skills, analyzed with the Pearson product-moment correlation coefficient.

Table 5

*Correlation Matrix between the Musical Sound Sensitivity Tests, English Prosody Processing Tests, and English Listening Comprehension Test in Study 1*

| Variable | Prosody input | Focus input | Chunking input | Prosody discrimination | Eiken Grade 3 listening test | Tonal test | Rhythm test |
|---|---|---|---|---|---|---|---|
| Prosody input | 1 | | | | | | |
| Focus input | .59*** | 1 | | | | | |
| Chunking input | .42*** | .44*** | 1 | | | | |
| Prosody discrimination | .17 | .09 | -.01 | 1 | | | |
| Eiken Grade 3 listening test | .40*** | .50*** | .35*** | .09 | 1 | | |
| Tonal test | .38*** | .37*** | .20 | .22 | .18 | 1 | |
| Rhythm test | .37*** | .47*** | .26* | .38*** | .31** | .78*** | 1 |

*Note:* *$p < .05$, **$p < .01$, ***$p < .001$.

Pearson's correlation analysis revealed the existence of a mid-level correlation between musical tone/rhythm processing skill and English prosody processing skills. Tone processing skill significantly correlated with prosody input ($r = .38$, $p < .001$) and focus input ($r = .37$, $p < .001$). Rhythm processing skill was found to have a significant correlation with (1) focus input ($r = .47$, $p < .001$), (2) prosody discrimination skill ($r = .38$, $p < .001$), (3) prosody input ($r = .37$, $p < .001$), and (4) chunking input ($r = .26$, $p < .05$). Considering that the distribution of the focus input, chunking input, tone, and rhythm data was distorted, in which skewness and kurtosis are high, Spearman's correlation analysis was also conducted, and the results are described comparatively with the results of the Pearson's analysis (see Appendix J).

Mid-level correlations were observed between Eiken Grade 3 listening test scores and some English prosody perception skills: (1) Eiken Grade 3 listening test score and focus input ($r = .50$, $p < .001$), (2) Eiken Grade 3 listening test score and prosody input ($r = .40$, $p < .001$), and (3) Eiken Grade 3 listening test score and chunking input ($r = .35$, $p < .001$). Mid-level correlation was also observed between Eiken Grade 3 listening test score and musical rhythm processing skill ($r = .31$, $p < .01$).

**3.5.3 Significant results of multiple regression analyses on the relationship between sound sensitivity variables, English prosody processing variables, and an English listening comprehension variable.** A multiple regression analysis was conducted to see whether sound sensitivity (Tonal test and Rhythm test) explains the variance in English prosody processing test scores and Eiken Grade 3 listening test scores, respectively. Tables 6–13 show the significant results of multiple regression analysis, with English prosody processing test scores and Eiken Grade 3 listening test scores as dependent variables, and the Sound sensitivity test scores as independent variables. Tables 14 and 15 show the significant results of multiple regression analysis, with Eiken Grade 3 listening test scores as dependent variables and English Prosody perception test scores as independent variables. This multiple regression analysis was conducted to see to what extent English Prosody perception test scores predict Eiken Grade 3 listening test scores.

Table 6

*Stepwise Multiple Regression Analysis with Prosody Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Ɓ* |
|---|---|---|---|
| Tone | .009 | .003 | .38** |
| Intercept | .61 | .085 | |

*Note:* $R^2$ = .14, ***p* < .005 (*p* = .001). Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$). B, *SEB, and β* are referred to in the Glossary.

Table 7

*Collective (Direct) Multiple Regression Analysis with Focus Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Ɓ* |
|---|---|---|---|
| Tone | -9.194E-006 | .003 | -.001 |
| Rhythm | .009 | .003 | .47** |
| Intercept | .64 | .065 | |

*Note:* $R^2$ = .22, ***p* < .01 (*p* = .006).

Table 8

*Stepwise Multiple Regression Analysis with Focus Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Ɓ* |
|---|---|---|---|
| Rhythm | .009 | .003 | .47*** |
| Intercept | .64 | .065 | |

*Note:* $R^2$ = .22, ****p* < .001 (*p* = .000). Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

Table 9

*Stepwise Multiple Regression Analysis with Chunking Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Rhythm | .005 | .002 | .26* |
| Intercept | .52 | .078 | |

*Note:* $R^2 = .068$, $*p < .05$ ($p = .024$). Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

Table 10

*Collective Multiple Regression Analysis with Prosody Discrimination Test as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| | B | *SEB* | *Β* |
|---|---|---|---|
| Tone | -.004 | .004 | -.17 |
| Rhythm | .013 | .004 | .51** |
| Intercept | .34 | .091 | |

*Note:* $R^2 = .15$, $**p < .01$.

Table 11

*Stepwise Multiple Regression Analysis with Prosody Discrimination as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Rhythm | .009 | .003 | .38** |
| Intercept | .38 | .089 | |

*Note:* $R^2 = .14$, $**p < .01$ ($p = .001$). Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

Table 12

*Collective Multiple Regression Analysis with Eiken Grade 3 Listening Test Score as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| | B | *SEB* | *Ɓ* |
|---|---|---|---|
| Tone | -.073 | .091 | -.14 |
| Rhythm | .24 | .10 | .42* |
| Intercept | 6.56 | 2.17 | |

*Note:* $R^2$ = .10, *$p$ < .05 ($p$ = .022).

Table 13

*Stepwise Multiple Regression Analysis with Eiken Grade 3 Listening Test Score as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Ɓ* |
|---|---|---|---|
| Rhythm | .18 | .065 | .31** |
| Intercept | 6.20 | 2.13 | |

*Note:* $R^2$ = .095, **$p$ < .01 ($p$ = .007). Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

Table 14

*Collective Multiple Regression Analysis with Eiken Grade 3 Listening Test Score as a Dependent Variable and Prosodic Perception as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Ɓ* |
|---|---|---|---|
| Prosody input | 2.58 | 2.88 | .12 |
| Focus input | 11.45 | 4.05 | .37** |
| Chunking input | 3.81 | 3.21 | .14 |
| Prosody discrimination test | .91 | 2.39 | .04 |
| Intercept | -4.22 | 3.33 | |

*Note:* $R^2$ = .28, **$p$ < .01.

Table 15

*Stepwise Multiple Regression Analysis with Eiken Grade 3 Listening Test Score as a Dependent Variable and Prosodic Perception as an Independent Variable in Study 1*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Focus input | 15.59 | 3.14 | .50*** |
| Intercept | -2.49 | 2.94 | |

*Note:* $R^2$ = .25, ****p* < .001 (*p* = .000). Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

Tables 7 and 8 show that sound sensitivity scores accounted for 22% of the focus input, where rhythm perception had a significant effect as identified from the final beta weights. Table 10 shows that rhythm perception had a significant effect in its accountability for 15% of prosodic discrimination. Tables 12 and 13 show the significant effect of rhythm perception on Eiken Grade 3 listening test scores. Tables 14 and 15 show the significant effect of focus input in its accountability for Eiken Grade 3 listening test scores.

**3.5.4 SEM analysis on the relationship between sound sensitivity variables, English prosody processing variables, and an English listening comprehension variable.** Structural equation modeling (SEM) was used to investigate how sound sensitivity, English prosody processing skill, and English listening comprehension causally affect each other. SEM reveals whether variables are interrelated through a set of linear relationships, by examining the variances and covariances of the variables. The SEM model shows that the variables are inter-related, with the path diagram in Figure 9. The IBM SPSS AMOS version 22 package released in 2015 was used for this investigation. The SEM model had good fit (*n* = 75, $\chi^2(13)$ = 10.29, *p* = .67, *ns.*; NFI = .94; CFI = 1.00; RMSEA = .00). The two variables of sound sensitivity presented high factor loadings with $\beta$ = 0.79 ($p$ < .001) for tone and $\beta$ = 0.98 ($p$ < .001) for rhythm, which means that rhythm represents the sound sensitivity variable remarkably highly. The four observed variables of English prosody processing skill presented various factor loadings from the lowest ($\beta$ = 0.20, $p$ > .05 (*ns*)) for prosody discrimination to the highest ($\beta$ = 0.83, $p$ < .001) for focus input. Causal relationships loading from sound sensitivity showed a significant effect to English prosody processing ($\beta$ = 0.57, $p$ < .001). Likewise, English prosody processing skill significantly contributed to English listening comprehension ($\beta$ = 0.51, $p$ < .001). Thus, English prosody processing skill directly affected English listening comprehension, and sound sensitivity did so indirectly. Figure 9 shows the causal relationships between sound sensitivity, English prosody processing skills, and English listening comprehension.

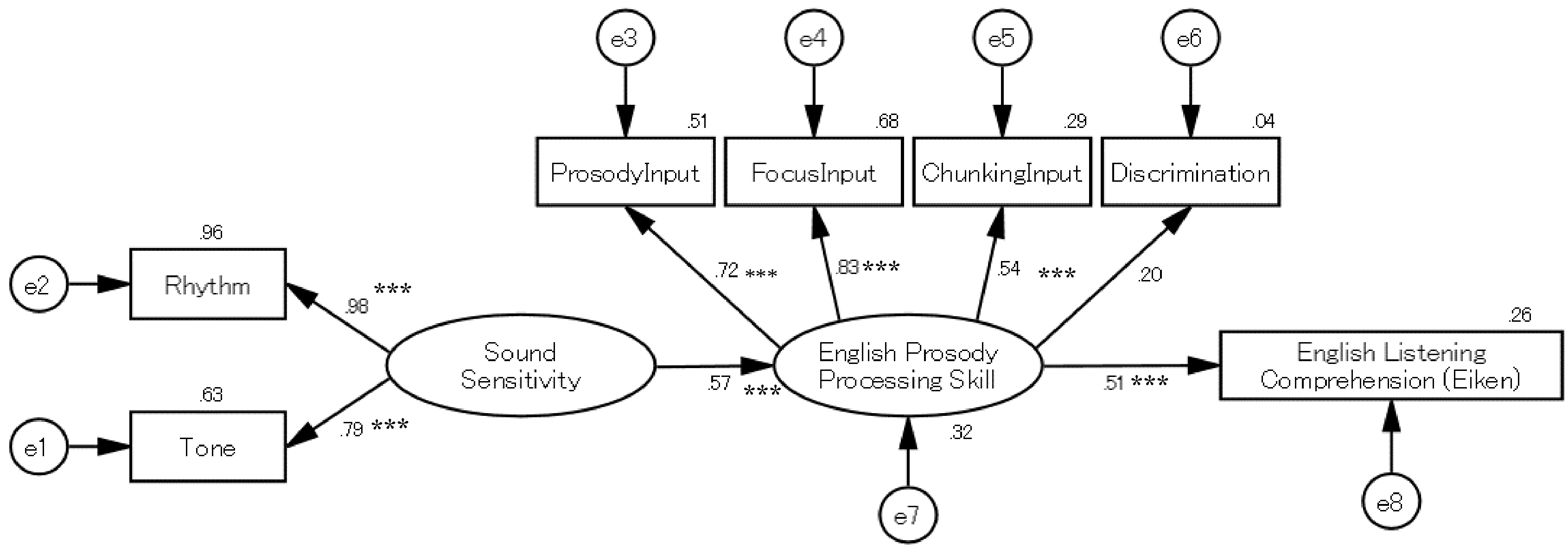



*Figure 9*. Causal relationships between sound sensitivity, English prosody processing, and English listening comprehension in Study 1. *Note:* $n = 75$; ***$p < .001$; the figures with asterisks indicate standardized regression coefficients. The figure on top of each box is the $R^2$.

Tables 16–19 show the fit indices for the model, unstandardized regression coefficients, standardized regression coefficients, and the statistical variance of measurement error.

Table 16

*Fit Indices for the Model in Study 1*

| CMIN | *Df* | *p* | CFI | TLI | RMSEA | LO90 | HI90 |
|---|---|---|---|---|---|---|---|
| 10.29 | 13 | .67 | 1.00 | 1.04 | .00 | .00 | .11 |

Table 17

*Unstandardized Regression Coefficients in Study 1*

| | Estimated value | *SE* | Test statistics | *p* |
|---|---|---|---|---|
| Prosody ← Sound sensitivity | .19 | .048 | 3.93 | *** |
| Tone ← Sound sensitivity | 1.00 | | | |
| Rhythm ← Sound sensitivity | 1.08 | .20 | 5.49 | *** |
| Prosody input ← Prosody | 1.00 | | | |
| Focus input ← Prosody | .82 | .15 | 5.43 | *** |
| Chunking input ← Prosody | .67 | .17 | 4.09 | *** |
| Discrimination ← Prosody | .19 | .13 | 1.52 | .13 |
| Eiken Grade 3 ← Prosody | .94 | .24 | 3.85 | *** |

*Note*: Eiken Grade 3 indicates Eiken Grade 3 listening test score.

Table 18

*Standardized Regression Coefficients in Study 1*

| | Estimated value |
|---|---|
| Prosody ← Sound sensitivity | .57 |
| Tone ← Sound sensitivity | .79 |
| Rhythm ← Sound sensitivity | .98 |
| Prosody input ← Prosody | .72 |
| Focus input ← Prosody | .83 |
| Chunking input ← Prosody | .54 |
| Discrimination ← Prosody | .20 |
| Eiken Grade 3 ← Prosody | .51 |

Table 19

*Statistical Variance of Measurement Error in Study 1*

| | Estimated value | *SE* | Test statistics | *p* |
|---|---|---|---|---|
| Sound sensitivity | 42.43 | 13.16 | 3.45 | *** |
| e1 | 26.61 | 8.44 | 3.15 | .002 |
| e2 | 2.42 | 8.44 | .29 | .78 |
| e3 | 4.88 | 1.10 | 4.44 | *** |
| e4 | 1.62 | .55 | 2.98 | .003 |
| e5 | 5.64 | 1.03 | 5.49 | *** |
| e6 | 4.68 | .78 | 6.03 | *** |
| e7 | 3.51 | 1.18 | 2.98 | .003 |
| e8 | 12.67 | 2.30 | 5.50 | *** |

## 3.6 Discussion

**3.6.1 The relationship between sound sensitivity and English prosody processing skills.** From the stepwise multiple regression analyses in Tables 6 to 11, rhythm perception had a significant effect on the accountability for Prosody discrimination test score ($\beta = .38$, $p < .01$), Focus input test score ($\beta$ = .47, $p$ < .001), Chunking input test score ($\beta$ = .26, $p$ < .05), and Eiken Grade 3 listening test scores ($\beta$ = .31, $p$ < .01). Rhythm test score did not explain Prosody input test score; rather, Tonal test score explained Prosody input test score ($\beta$ = .38, $p$ < .01).

The rhythm perception represented by Rhythm test scores makes it possible to process the musical rhythm composed of the durational patterns of sounds and pauses at all levels. As stated in 2.1.3, there is a difference between the syntactic rhythm of language and musical rhythm (Jackendoff & Lerdahl, 1983, p. 330); therefore, it cannot be simply concluded that listeners with effective musical rhythm processing skills would have effective syntactic rhythm for every language. However, musically sophisticated learners might process an unknown linguistic rhythm flexibly by activating various areas of their brain. Milovanov et al. (2009) showed that musically sophisticated listeners reacted more acutely to durational mismatch both in musical and language sounds, presenting negative reactions. This phenomenon might imply that musically sophisticated listeners are generally sensitive to both musical and linguistic sounds. Considering this significant correlation of rhythm processing skill with English prosody processing skill in Study 1, high rhythm sensitivity in music can contribute to better perception of English rhythmical prosody.

Firstly, the significant predictability of rhythm processing skill with the Prosody discrimination test ($R^2 = .14$, $\beta = .38$, $p < .01$) is discussed. The Prosody discrimination test in this study investigates the ability to distinguish English prosody from the prosody of another language. In this test, 12 pairs of English and another language have unique syntactic rhythms, so that rhythm sensitivity may play an important role in distinguishing between the pairs. Although statistical difference has not yet been investigated, the highest score ($M$ =.84) was obtained for the question that asked participants to distinguish English from Arabic. They are both stress-timed languages and, therefore, this unexpected result may be due to factors other than their rhythms, such as the speaking speed, pronunciation clarity, or expressiveness of the recorded speech prosodies. The second highest score ($M$ = .81) was for the question asking participants to distinguish English from Japanese. The remarkable difference between these prosody pairs and the participants' familiarity with Japanese prosody as their native language explain this result reasonably.

Secondly, the significant predictability of rhythm processing skill with the Focus input test ($R^2 = .22$, $\beta = .47$, $p < .001$) is discussed. In the Focus input test, from pairs of a word, it is seen that word prominence is placed differently in a phonetic structure. A prominent syllable is pronounced relatively stronger than the rest. According to Markus (2005), English has a rhythmic flow when spoken, and nuclear-stress placement is central to its phonological character.

Markus (2005) also wrote, “English tends to adapt its quantity of syllables to a rhythm conditioned by stresses” (p. 106). Another theory on English stress and rhythm is the “Rhythm Rule” (Liberman & Prince, 1977). Under the Rhythm Rule, the stress is shifted leftward in words such as thirteen and Tennessee, whose stresses are on the final syllables when spoken in isolation (Koine, 1995). The idea behind this rule is that, in terms of metrical structure, the stresses of a linguistic phrase form a hierarchical metrical pattern, with relatively heavy stresses corresponding to relatively strong beats and relatively weak stresses corresponding to relatively weak beats. Gimson (1962) says, “the rhythmic pattern of a word remains constant whatever the environment, retaining its rhythmic identity in the total rhythmic grouping of the longer utterance” (p. 265). This means that the relation between the primary and secondary accents in a word is not changed, although the primary accent on a word may lose its position as the nuclear accent of a phrase or sentence. However, there are exceptional cases where the internal stress relationships change, for example, when “thirtéen” becomes “thìrteen mén” or when “Tennessée” becomes “Tènnessee áir.” These internal stress changes prevent a “metrical clash,” that is, the condition where serious stress vowels do not have any interval between them, such as in the examples of “thirteen men” or “Tennessee air.” In other words, the Rhythm Rule avoids metrical clashes “by reversing the largest weak and strong of the first word, thereby separating the second-strongest stress from the strongest by a relatively weak stress" (Jackendoff & Lerdahl, 1983, p. 325). Therefore, the Rhythm Rule can be regarded as a rule that makes prosodic structure change to conform closely to an ideal metrical pattern. At the phrase level in English, its rhythmic metrical pattern with regular stresses follows the Rhythm Rule. The Focus input test asked if participants perceive the focused word with the heaviest available syllable, that is, the stressed word, in the spoken sentences. The lexical rhythmic unit of English is the foot, with each foot having a stress that makes lexical prominence (Jun, 2012, p. 444). Regular stresses and English rhythm are closely related to each other. Therefore, it is reasonable to infer that the Focus input test score relates to the participants' rhythm sensitivity.

Thirdly, the significant predictability of rhythm processing skill with the Chunking input test ($R^2 = .068$, $\beta = .26$, $p < .05$) is discussed. The Chunking input test assesses the comprehension of prosodic boundaries of phrases. There is a proposed "segmental anchoring" whereby the beginning and end of a linguistically significant pitch movement are anchored to specific locations in the segmental structure, which means that the slope and duration of the pitch movement vary according to the segmental material with which they are associated (Arvaniti, Ladd, & Mennen, 1998). The possible existence of segmental anchoring, whereby the duration of pitch movements in speech is adjusted to the duration of the accompanying segmental material

(Ladd, 2006), was empirically probed. Ladd, Dan, Hanneke, and Schepman (1999) showed that the alignment of the valley and peak of an English rising pitch accent is unaffected by the changes in segmental duration brought about by modifications in the speech rate. The duration and slope of the accents become shorter and steeper as speech rate increases. If this phenomenon of segmental anchoring is applied in English, the relative distances between accented syllables in accompanying segments of phrases or sentences become aligned, which creates a regular rhythm in the phrases or sentences. Considering such segmental anchoring, the Chunking input test may measure participants' perception of rhythm, composed of the duration of pitch movements and their segments in test phrases. From Pearson's correlation analysis, chunking input processing was significantly correlated with rhythm perception ($r = .26$, $p < .05$)

Fourthly, the significant predictability of the Prosody input test score with Tonal test score ($R^2 = .14$, $\beta = .38$, $p < .01$) is discussed: Words that receive nuclear stress have low pitch or high mounted pitch when pronounced. In addition to the prominent relationship between syllables, the pitch range shifts between such syllables: "the upward modification of the pitch range at the beginning of a new stretch of declination–and final lowering–the corresponding downward modification of the pitch range at the end" (Ladd, 2008, p. 308). Considering the Rhythm Rule, English is phonetically called a stress accent language, while Japanese is a non-stress accent language where the lexical accent is not marked by stress at all but only by pitch movement. In English, pitch features are only postlexical or intonational, while in Japanese, pitch features are specified in the lexicon (Ladd, 2008, p. 164). Tone sensitivity did not significantly predict the Focus input test, which may relate to the fact that English stress accent does not always coincide with pitch movement. In English, the lexical accent is regularly manifested through stress, including increased intensity and duration (Beckman & Pierrehumbert, 1986). Beckman and Pierrehumbert (1986) indicate that the typology of the dynamic accent (stress accent) and melodic accent (pitch accent) is valid, experimentally supporting the phenomenon that duration, intensity, and vowel quality all play a significant role in English, whereas pitch change is only an acoustic cue to the accent in Japanese. The Prosody input test helps to determine the ability to perceive prosodic differences where the participants identify whether the intonation or prosody of pairs of phrases is the same or different. The intonation tested herein is postlexical; it minimizes the features of stress, accent, and tone, which are determined in the lexicon. Postlexical intonation is created by pitch movements; therefore, this test can be regarded as the test for pitch contours. Fortunately, Tonal test scores significantly predicted Prosody input test scores in Study 1.

**3.6.2 The relationship between English prosody processing skills and English listening comprehension.** As a result of Pearson's correlation analysis, one of the Prosody perception test scores, the Focus input test score correlated with the Eiken grade 3 listening test score ($r$ = .50, $p$ < .001). Further, the results of the collective multiple regression analysis showed that English prosody perception skills explained the Eiken Grade 3 listening test scores by 28%, where focus input had a significant effect on accountability ($\beta$ = .37, $p$ < .01). This result implies that focusing perception significantly predicts English listening comprehension. From these results of Study 1, the importance of rhythm perception (see 3.6.1) and focus perception in the prosodic and semantic understanding of English was deduced.

English is a stress-timed language, whose regularly arising stresses make sentences sound rhythmic. In contrast, Japanese is often called a mora-timed language, which is a variation of a syllable-timed language (see Chapter 2). Thus, Japanese utterances sound longer and flatter in general (as described in Chapter 2). As English rhythm perception and production have been emphasized for enhancing native-like pronunciation (Anderson-Hsieh et al., 1992), acquisition of English prosody may need to be further emphasized in the Japanese EFL environment, especially in the primary years of compulsory education. Discussion on the best period for starting EFL education will be made in Chapter 7.

**3.6.3 The relationship between sound sensitivity and English listening comprehension.** One of the sound sensitivity variables, i.e., rhythm processing skill, significantly correlated with the Eiken Grade 3 listening test score ($r = .31$, $p < .01$). Tables 12 and 13 of the multiple regression analysis show the significant accountability of Eiken Grade 3 listening test scores by rhythm processing skill ($R^2 = .10$, $\beta = .42$, $p < .05$; $R^2 = .095$, $\beta = .31$, $p < .01$). These results from the correlation analysis and multiple regression analysis imply that rhythm processing skill might directly contribute to English listening comprehension.

The SEM analysis result increased the probability that sound sensitivity indirectly contributed to the English listening comprehension of the high school participants. According to the SEM analysis, the two variables of sound sensitivity presented high factor loadings with $\beta = .79$ ($p < .001$) for tone and $\beta = .98$ ($p < .001$) for rhythm, which means that tone and rhythm highly represent sound sensitivity as a latent factor. The second latent factor, i.e., English prosody processing skill, had causal relationships with sound sensitivity. The causal relationships loading from sound sensitivity showed a significant effect on English prosody processing ($\beta = .57$, $p < .001$). Likewise, English prosody processing skill significantly contributed to English listening comprehension ($\beta = .51$, $p < .001$). The SEM analysis affirmed the model composed of sound sensitivity, English prosody processing skills, and English comprehension skill. The model implies that English prosody processing skill directly affected English listening comprehension, and sound sensitivity indirectly affected English listening comprehension (see Figure 9 in section 3.5.4).

Looking over the results of Study 1, the research questions (1) “For high school students, does sound processing skill contribute to EFL prosody perception?” (2) “With regard to high school students, does sound sensitivity contribute to EFL listening comprehension?” (3) “With regard to high school students, does English prosody perception contribute to English listening comprehension in EFL?” were partly answered in the affirmative.

In further research, the English prosody measurement PEPS-C needs to be pedagogically used as a measurement of testees’ prosodic proficiency, not merely by screening them into normal/impaired or acceptable/inacceptable. PEPS-C has been developed and used for both educational and clinical purposes. Prosody processing skill in EFL should be assessed for various purposes, in a similar way as language aptitude tests had been developed for political and educational purposes.

In the next chapter, Study 2 will be conducted to clarify how sound processing skill, particularly tone/rhythm perception skill and rhythm production skill, the receptive/productive English prosody processing skills, and English listening comprehension of Japanese elementary school children are related to each other. These child participants have been taking school English classes twice a week for nearly three years; thus, they are at an earlier stage of FLA, compared to the participants in Study 1. Moreover, other data, in addition to the data from the perception tests used for Study 1, will be collected concerning participants' productive English prosody and productive musical rhythm.

# Chapter 4 Study 2: Relationship between Sound Sensitivity, Receptive/Productive Prosody Processing of English, and English Listening Comprehension of Elementary School Children

## 4.1 Purpose of Study 2

From the experimental study in Chapter 3, the correlation between sound sensitivity, English prosody processing skill, and English listening comprehension was implied for high school participants. As a parallel study, the same sound sensitivity perception and English prosody processing measures as used in Study 1, in addition to a supplemental Prosody production test, a Rhythm production test, and the Junior English Test (JET) as an English listening comprehension test, were conducted on elementary school participants with approximately three years of EFL experience. This parallel and supplemental study was conducted to ascertain the correlation between sound perception/production, English prosody perception/production, and English listening comprehension, in third-year elementary school participants. For Study 2, the English prosody production test and Rhythm production tests were added to the test materials used in Study 1, in order to investigate if both perceptive and productive rhythm processing skills correlate with perceptive/productive English prosody processing skills. A Tone production test was not added lest elementary school participants should be too nervous to reproduce melodies in front of an instructor. Furthermore, as an English listening test, the JET grade 7–8 (see Appendix M for details) was used, instead of Eiken Grade 3 listening test, as JET is suitable for any age group with one–three years of English learning experience and was developed to test listening comprehension skill. Thus, the research questions to be answered in Study 2 are as follows: (1) For elementary school students, does sound perception/production contribute to EFL prosody perception/production?; (2) For elementary school students, does sound sensitivity contribute to English listening comprehension?; (3) For elementary school students, does English prosody perception/production contribute to English listening comprehension?

**4.2 Participants**

Fifty-three students at an elementary school cooperated to participate in the experiments. They were at the end of their third year of elementary school at the beginning of the study, and in the middle of their fourth year when the study was completed. In their first and second years, an American native speaker taught them English communication for two hours per week. In their third and fourth years, a native teacher and a Japanese teacher taught one class each, with a total of two classes per week. The 53 students had mostly similar socioeconomic backgrounds. At the time of the study, 29 students had been taking private English lessons outside of school, 22 of whom had begun taking lessons more than three years prior to the study, and the remaining 7 of whom had begun taking private lessons within three years of the study. The remaining 24 students only study English at school. The English education background of the 53 participants was not controlled, which will make it difficult to consider years of learning as an independent variable explaining any dependent variables.

**4.3 Materials**

1. PEPS-C (Peppé et al., 2007): The content of this test is described in 3.3, and Appendices A1 and A2.

2. Prosody discrimination test: The content of this test is described in 3.3, and in Appendices A3, A4, and E.

3. Prosody production test

A subtest of PEPS-C, “Prosody output” (Peppé et al., 2007), which assesses the ability to imitate prosodic forms, was used. The participants reproduce pronounced phrases such as “green-and-red and black socks,” “cream-buns and chocolate,” and “red and blue socks.” See Appendix K for the test script.

4. Measurement of Musical Aptitude: The content of this test is described in 3.3, and Appendices

C and D.

5. Rhythm production test

This test was created for an experimental purpose by the author to measure the ability to imitate musical rhythm. It contains 12 rhythmic beats by four measures in 2/4, 6/8, 3/4, 4/4, and 12/8 time, which are played with a melodica by the examiner and reproduced with handclaps by the students. These rhythms were chosen from Mother Goose, the traditional British nursery rhyme. See Appendix K for the scoring sheet and Appendix L for the notes used for this test.

6. JET 7–8 level: Measurement of English listening and reading

The JET consists of 30 multiple-choice questions to test the English communication skill of English learners with approximately one to three years of learning experience. Thirty-five minutes were allotted to answer the 30 test questions, including 25 listening questions and 5 reading/vocabulary questions. The maximum score is 30 points in this study. The listening section consists of six parts: parts 1–2 test vocabulary meanings; parts 3–4 test sentence meanings; part 5 tests conversation meaning; and part 6 tests communication skill. The reading section tests written word meanings. See Appendix M for the script.

7. Questionnaire 1: This questionnaire contains questions on years of music training, instruments or areas of training, year of starting music training, training hours, motivation to train, and ability

to read musical notation. See Appendix N for the script.

8. Questionnaire 2: This questionnaire contains questions on years of EFL training at home and at school. The other questions concern studying hours, motivation, interest in EFL, and degree of receptive use of English such as listening and reading. See Appendix O for the script.

**4.4 Procedures**

All procedures of the experiment and data analysis are listed below:

1. On March 5, 2014, the two major elements of musical aptitude (tonal and rhythmic abilities) were measured by administering the Intermediate Measures of Music Audiation test (Gordon, 1982) to 53 students in their two homeroom classes.

2. On June 4, 2014, the English prosody processing skills of participants were measured with PEPS-C, and English prosody discrimination skill was measured with the Prosody discrimination test. The students took these tests in their two homeroom classes.

3. On July 4, 2014, the students completed the questionnaires on their music and English learning in their two homeroom classes.

4. From May to July 2014, the Rhythm production test and Prosody production test were administered to the students. These tests were administered individually to participants in a separate room at school, either during lunch break or after school. These productive tests took

each student 15–20 minutes to complete.

5. In January 2015, the JET 7–8 level was administered to students during their English classes.

## 4.5 Results

The analysis and statistical results of the experiments on 53 private elementary school students in the third to fourth year will be given as (1) descriptive statistics, (2) Pearson’s correlation, (3) stepwise and collective multiple regression analyses, and (4) SEM analysis.

**4.5.1 Descriptive statistics.** Table 20 shows the descriptive statistics for all the tests in this study. Inter-rater coefficients on the data of the productive measure will be given after the descriptive statistics. The result of a *t*-test on gender difference in the musical aptitude of the participants will also be described as supplemental data.

Table 20

*Descriptive Statistics for the Tests Used in Study 2*

| Variable | *K* | *M* | *SD* | Range | *α* | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| **Prosodic skill measurements:** | | | | | | | |
| Focus input | 16 | 14.96 | 1.18 | 12-16 | .37 | -1.10 | .45 |
| Chunking input | 18 | 11.11 | 2.61 | 6-18 | .41 | .63 | -.068 |
| Prosody input | 16 | 14.21 | 2.28 | 7-16 | .75 | -1.57 | 1.87 |
| Prosody discrimination | 12 | 6.38 | 2.19 | 2-12 | .40 | -.29 | -.71 |
| Prosody production | 16 | 6.80 | 4.12 | 0-16 | .89 | .69 | -.37 |
| **English listening skill measurement:** | | | | | | | |
| JET | 30 | 22.1 | 4.60 | 12-30 | .80 | -.023 | -.87 |
| **Musical aptitude measurement:** | | | | | | | |
| Tone perception | 40 | 35.04 | 3.72 | 22-40 | .77 | -1.41 | 2.66 |
| Rhythm perception | 40 | 31.60 | 4.25 | 14-38 | .68 | -1.51 | 4.34 |
| Rhythm production | 12 | 7.34 | 1.93 | 2-11 | .78 | -.36 | -.31 |

*Note: K* indicates the number of items in each test.

Table 20 shows the descriptive statistics for all tests in this study, including mean, standard deviation, minimum score, maximum score, and test reliability (Cronbach's *α*). The test reliability of the Prosody discrimination tests was rather low (*α* = .40). Cronbach's *α* was smaller on the Focus input test (*α* = .37) and Chunking input test (*α* = .41), compared to that of the high school participants' focus input (*α* = .83) and chunking input (*α* = .60). This low reliability of

Focus input test and Chunking input test when administered to elementary school participants was possibly caused by the small sample size or variability, which requires improvement in future research. The focus data may not have had sufficient statistical variability, due to the ceiling effect in the case of elementary school participants. The statistical reliability and validity of the Focus input test and Chunking input test have been guaranteed as the published test materials for clinical and educational use for PEPS-C (Peppé & McCann, 2003). However, the Prosody discrimination test was originally produced by the author for this study and 12 questions may not be sufficient to ensure its reliability. To increase the low reliability of prosody discrimination, the test questions with the lowest four scores were trimmed from the test and analyzed: When questions 1, 2, 10, and 16 were trimmed, the Cronbach's $\alpha$ rose to .61 (Appendix P). When Pearson's correlation analysis and the multiple regression analysis were conducted with the trimmed version of the Prosody discrimination test, the result of Pearson's correlation analyses on the trimmed data showed a weaker correlation with Prosody production ($r$ = .26, *ns*) and Rhythm production ($r$ = .20, *ns*). The results of the multiple regression analyses on the trimmed and raw data had no statistically noticeable differences (see Appendix P for the results). These comparative analyses between the raw and trimmed data mean that the raw data of the prosody discrimination should be mostly reliable and valid for analysis with Pearson's correlation analysis and multiple regression analysis, although the weaker correlation with Prosody production ($r = .26$, *ns*) and Rhythm production ($r = .20$, *ns*) should be taken into account when discussing the results.

The inter-rater coefficient of the Prosody production test was calculated between the two raters' evaluation scores. Two raters evaluated the recording under the criteria.[17] One rater was Japanese and had been teaching English at high schools for 13 years. The other rater was a native speaker of American English who had been teaching at elementary and junior high schools as an assistant English teacher for four years in Japan. The inter-rater reliability of prosody production was high ($\alpha = .98$). A table of inter-rater reliability scores can be seen in Appendix Q (Table Q1).

The inter-rater coefficient of the Rhythm production test was calculated between the two raters' evaluation scores. Two raters evaluated the recording using the criteria.[18] The raters of the Rhythm production test were two doctoral students from a national university. One rater had been taking violin lessons for six years and practiced the violin for 30 minutes per day. The other rater was a professional pianist and an assistant professor in the music department of an arts university, and had given private piano lessons at home. The inter-rater reliability of the Rhythm production test was also high ($\alpha = .96$). A table of inter-rater reliability scores can be seen in Appendix R (Table R1). Lest there be any difference in the sound processing skill of male and female elementary school students, an independent *t*-test was conducted on their tonal and rhythm scores. The data are shown in Appendix S for the tone scores and Appendix T for the rhythm scores. According to the results of the *t*-test, no significant difference was found between the average tone scores of 27 males ($M$ = 35.5) and 26 females ($M$ = 34.5) ($t$ (51) = .90, $p$ = .37). No significant difference was found between the average rhythm scores of 27 males ($M$ = 31.5) and 25 females ($M$ = 31.9) ($t$ (51) = -.37, $p$ = .71).

**4.5.2 Pearson's correlation analysis between sound perception/production variables, English prosody processing variables, and a listening comprehension variable.** A correlation analysis was conducted between musical sound processing tests and the English prosody test scores, and the results are shown in Table 21.

Table 21

*Pearson's Correlation Matrix between the Musical Sound Perception/Production Tests and the English Prosody Tests in Study 2*

| Variable | Prosody input | Focus input | Chunking input | Prosody discrimination | Prosody production | Tone | Rhythm | Productive Rhythm | JET |
|---|---|---|---|---|---|---|---|---|---|
| Prosody input | 1 | | | | | | | | |
| Focus input | .21 | 1 | | | | | | | |
| Chunking input | .29* | .18 | 1 | | | | | | |
| Prosody Discrimination | .01 | .27 | .15 | 1 | | | | | |
| Prosody production | .30* | .19 | .42** | .31* | 1 | | | | |
| Tone | .54*** | .02 | .15 | .28* | .35* | 1 | | | |
| Rhythm | .51*** | .18 | .001 | -.058 | .042 | .46*** | 1 | | |
| Productive Rhythm | .17 | .38** | .36** | .36** | .20 | .20 | .24 | 1 | |
| JET | .36** | .14 | .49*** | .036 | .72*** | .18 | .13 | .027 | 1 |

*Note:* **p* < .05, ***p* < .01, ****p* < .001.

As shown in Table 21, significant correlations were found between the three subtests of sound perception/production and the five subtests of English prosody perception/production. Firstly, the Tonal test was significantly correlated with prosody input ($r = .54$, $p < .001$), prosody production ($r = .35$, $p < .05$), and prosody discrimination ($r = .28$, $p < .05$). Secondly, the Rhythm test correlated only with prosody input ($r = .51$, $p < .001$). Thirdly, rhythm production correlated with focus input ($r = .38$, $p < .01$), chunking input ($r = .36$, $p < .01$), and prosody discrimination ($r = .36$, $p < .01$). A significant correlation can also be found between JET and

English prosody perception/production: JET scores correlated with prosody input ($r = .36$, $p < .01$), chunking input ($r = .49$, $p < .001$), and prosody production ($r = .72$, $p < .001$). These results of the correlation analyses show that there are some correlations (1) between several English prosody processing skills and either or both sound perception/production ability, and (2) between English prosody processing skills and English listening comprehension skill. Although some of the correlations are not very strong between these variables, it is necessary to ascertain how these variables are correlated. A precise discussion will be provided in the discussion section to follow.

**4.5.3 Significant results of multiple regression analyses on the relation between sound perception/production variables, English prosody processing variables, and an English listening comprehension variable.** A stepwise multiple regression analysis and collective multiple regression analysis were conducted to ascertain to what extent each score in the musical Sound sensitivity test scores accounts for each English prosody processing test score. From some collective analyses, it was found that the assumed model, which assumes that every independent variable explains a dependent variable, did not fit well. In such cases, a stepwise multiple regression analysis was conducted using only the high factor-loading independent variable. Tables 22 to 31 show the significant results of the collective or stepwise multiple regression analysis, conducted with English prosody processing test scores as dependent variables, and Sound sensitivity test scores as independent variables. Tables 32–33 show the significant results of the collective and stepwise multiple regression analysis with the JET score as a dependent variable and English prosody perception/production as independent variables.

Table 22

*Collective Multiple Regression Analysis with Prosody Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Tone | .24 | .078 | .39** |
| Rhythm | .17 | .069 | .32* |
| Rhythm production | .019 | .14 | .016 |
| Intercept | .20 | 2.63 | |

*Note:* $R^2$ = .38, **p* < .05, ***p* < .01.

Table 23

*Stepwise Multiple Regression Analysis with Prosody Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Tone | .24 | .077 | .39** |
| Rhythm | .18 | .067 | .33* |
| Intercept | .25 | 2.58 | |

*Note:* $R^2$ = .38, ***p* < .001 (*p* = .003), *p** < .05 (*p* = .012). Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

Table 24

*Collective Multiple Regression Analysis with Focus Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Tone | -.037 | .047 | -.12 |
| Rhythm | .041 | .041 | .15 |
| Rhythm production | .22 | .083 | .36** |
| Intercept | 13.36 | 1.57 | |

*Note:* $R^2$ = .16, ***p* < .01.

Table 25

*Stepwise Multiple Regression Analysis with Focus Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Rhythm production | .23 | .079 | .38** |
| Intercept | 13.26 | .60 | |

*Note:* $R^2$ = .14, ***p* < .01 (*p* = .005); Stepwise selection (selection criteria: stepwise only selects independent variables with *p* ≤ .05).

Table 26

*Collective Multiple Regression Analysis with Chunking Input as a Dependent Variable and Sound Processing as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Tone | .11 | .10 | .15 |
| Rhythm | -.10 | .09 | -.16 |
| Rhythm production | .49 | .19 | .36* |
| Intercept | 6.86 | 3.51 | |

*Note:* $R^2$ = .15, **p* < .05.

Table 27

*Stepwise Multiple Regression Analysis with Chunking Input as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *ℬ* |
|---|---|---|---|
| Rhythm production | .48 | .18 | .36** |
| Intercept | 7.56 | 1.35 | |

*Note:* $R^2$ = .13, ***p* < .01 (*p* = .009); Stepwise selection (selection criteria: stepwise only selects independent variables with *p* ≤ .05).

Table 28

*Collective Multiple Regression Analysis with Prosody Discrimination as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *ℬ* |
|---|---|---|---|
| Tonal test | .20 | .082 | .34* |
| Rhythm | -.16 | .073 | -.30* |
| Rhythm production | .41 | .15 | .37** |
| Intercept | 1.34 | 2.77 | |

*Note:* $R^2$ = .24, **p* < .05, ***p* < .01.

Table 29

*Stepwise Multiple Regression Analysis with Prosody Discrimination as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *ℬ* |
|---|---|---|---|
| Rhythm production | .36 | .15 | .33* |
| Intercept | 3.77 | 1.12 | |

*Note:* $R^2$ = .11, **p* < .05 (*p* = .02); Stepwise selection (selection criteria: stepwise only selects independent variables with *p* ≤ .05).

Table 30

*Collective Multiple Regression Analysis with Prosody Production as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Tonal test | .44 | .16 | .40** |
| Rhythm test | -.17 | .14 | -.18 |
| Rhythm production | .35 | .29 | .16 |
| Intercept | -5.75 | 5.48 | |

*Note:* $R^2$ ＝ .17, **p* < .05, ***p* < .01.

Table 31

*Stepwise Multiple Regression Analysis with Prosody Production as a Dependent Variable and Sound Sensitivity as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Tonal test | .37 | .15 | .34* |
| Intercept | -6.15 | 5.11 | |

*Note:* $R^2$ ＝ .12, **p* < .05 (*p* = .013); Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

Table 32

*Collective Multiple Regression Analysis with JET Score as a Dependent Variable and English Prosody Perception/Production as Independent Variables in Study 2*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Prosody input | .26 | .20 | .13 |
| Focus input | .056 | .39 | .014 |
| Chunking input | .35 | .18 | .20 |
| Prosody discrimination | -.36 | .21 | -.17 |
| Prosody production | .72 | .12 | .64*** |
| Intercept | 11.01 | 5.70 | |

*Note:* $R^2$ ＝ .60, ****p* < .001.

Table 33

*Stepwise Multiple Regression Analysis with JET Score as a Dependent Variable and English Prosody Perception as an Independent Variable in Study 2*

| Predictor | B | *SEB* | *Ɓ* |
|---|---|---|---|
| Chunking input | .39 | .18 | .22* |
| Prosody production | .70 | .12 | .63*** |
| Intercept | 12.96 | 1.90 | |

*Note:* $R^2$ = .56, ****p* < .001 (*p* = .000), **p* < .05 (*p* = .04); Stepwise selection (selection criteria: stepwise only selects independent variables with $p \leq .05$).

The results of the multiple regression analysis demonstrated that either or both tone/rhythm perception or rhythm production significantly explained English prosodic perception/production. Tables 22–33 show that (1) the Tonal test had a significant effect in accounting for prosody input, prosody discrimination, and prosody production, (2) the Rhythm test had a significant effect in accounting for prosody input and prosody discrimination, (3) the Rhythm production test had a significant effect in accounting for focus input, chunking input, and prosody discrimination. Table 33 shows that chunking input had a significant effect in accounting for JET score among prosody perception variables. It also shows that prosody production was the most significant variable in explaining JET scores among the prosody perception/production variables.

A numerical explanation of the results is as follows: The collective multiple regression analysis was conducted with tone perception, rhythm perception, and rhythm production as predictors for prosody input, focus input, chunking input, prosody discrimination, and prosody production. When necessary, a stepwise multiple regression analysis was additionally conducted to obtain higher predictability with influential independent variables. As shown in Table 22, sound sensitivity, consisting of tone perception, rhythm perception, and rhythm production, accounted for almost 38% of the variance in prosody input, with tone perception being identified as a significant predictor ($\beta = .39$, $p < .01$) in addition to rhythm perception ($\beta = .32$, $p < .05$). Table 23 shows that sound sensitivity accounted for almost 38% of the variance in prosody input with tone perception ($\beta = .39$, $p < .001$) and rhythm perception ($\beta = .33$, $p < .01$) represented as significant predictors. As shown in Table 24, sound sensitivity accounted for almost 16% of the variance in focus input, with rhythm production identified as a significant predictor ($\beta = .36$, $p < .01$). Table 25 shows that rhythm production accounted for almost 14% of the variance in focus input as a significant predictor ($\beta = .38$, $p < .01$). As shown in Table 26, sound sensitivity accounted for almost 15% of the variance in chunking input, with rhythm production identified as a significant predictor ($\beta = .36$, $p < .05$). Table 27 shows that rhythm production accounted for almost 13% of the variance in chunking input as a significant predictor ($\beta = .36$, $p < .01$). As shown in Table 28, sound sensitivity accounted for almost 24% of the variance in prosody discrimination with tone perception ($\beta = .34$, $p < .05$), rhythm perception ($\beta = -.30$, $p < .05$), and rhythm production ($\beta = .37$, $p < .001$) as significant predictors. As shown in Table 29, rhythm production accounted for prosody discrimination at almost 11% as a significant predictor ($\beta = .33$, $p < .05$). As shown in Table 30, sound sensitivity accounted for almost 17% of the variance in prosody production, with tone perception identified as a significant predictor ($\beta = .40$, $p < .01$). Table 31 shows that tone perception uniquely accounted for prosody production at almost 12% as the most significant predictor ($\beta = .34$, $p < .05$).

Tables 32 and 33 show the relationship between sound sensitivity with JET score, based on corrective multiple regression analysis and stepwise multiple regression analysis. Table 32 shows that English prosody processing accounted for almost 60% of the variance in JET score, with prosody production as a significant predictor ($\beta = .64$, $p < .001$). Table 33 shows that chunking input ($\beta = .22$, $p < .05$) and prosody production processing ($\beta = .63$, $p < .001$) accounted for almost 56% of the variance in JET score as significant predictors.

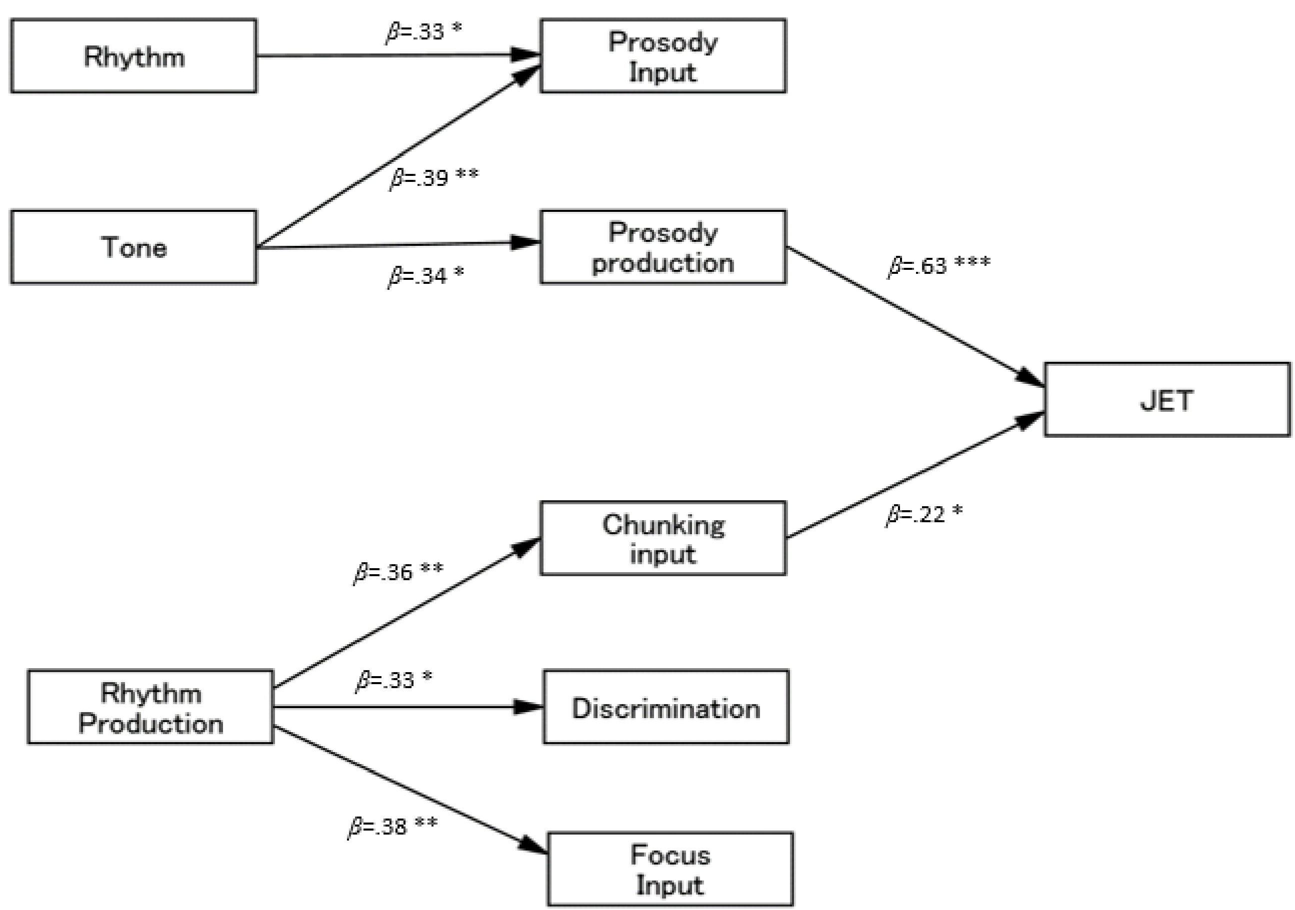



*Figure 10.* Causal relationships between sound sensitivity, English prosody processing, and English listening comprehension in Study 2. *Note:* $n = 53$; $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; the figures indicate standardized regression coefficients.

SEM was carried out to investigate how much sound sensitivity, English prosody processing skill, and English listening comprehension causally affect each other. The SEM analysis was conducted to investigate the causal model by those three latent variables. However, the fit indices for the model did not show reasonable fit for explaining the relationship between the variables ($n$ = 53, $\chi^2(26)$ = 70.93, $p$ = .000; NFI = .54; CFI = .62; RMSEA = .18). Significant relationships were found between the variables or the latent variables, so that they are worth re-examining. Future research is required to explore a better model.

## 4.6 Discussion

**4.6.1 The relationship between sound sensitivity and English prosody processing skills.** The results of Pearson's correlation analysis in Table 21 showed that the Tonal test scores correlated with prosody input and prosody discrimination, the Rhythm test correlated with prosody input, and the Rhythm production test correlated with prosody input, chunking input, and prosody discrimination. Each correlation found among these variables will be discussed below.

The correlation between Tonal test scores and Prosody input test scores ($r$ = .54, $p$ < .001) has already been discussed in 3.6, where it was reasoned that in this correlation, the pitch movement perception was tested in both the Prosody input test and the Tonal test. Prosody input test scores also correlated with Prosody production test scores ($r = .35$, $p < .05$), which may be explained by the fact that both tests measure intonation processing skills, even if one is asking for their perception and the other for their production.

As shown in Table 30, sound sensitivity accounted for almost 17% of the variance in prosody production, with tone processing identified as a significant predictor ($\beta = .40$, $p < .01$). Table 31 shows that tone perception uniquely accounted for prosody production at almost 12% as the most significant predictor. These results also imply that pitch perception contributes to English prosody production. English is regarded as a dynamic accent language, i.e., a stress accent language with postlexical pitch, that is, intonation. In English, pitch features are not specified at a word level but rather at a sentence level (Beckman & Pierrehumbert, 1986). English is not a pitch accent language as Japanese is; however, pitch perception might be essential in rehearsing the spoken speech in the Prosody production test.

A mid-level correlation was found between the Rhythm test and Prosody input test ($r = .51$, $p < .001$). The Prosody input test asks for rhythm and intonation that are created with intensity, duration, and controlled pitch. The prosodies in the Prosody input test were the low pass filtered sound signals of English phrases, where features of intensity, duration, and pitch (0–

300Hz) remained. Thus, it is reasonable that the Prosody input test scores correlated with both the Tonal test scores and Rhythm test scores. As reported in the results of stepwise multiple regression analysis shown in Table 23, sound sensitivity accounted for 38% of the variance in prosody input with tone perception ($\beta = .39$, $p < .001$) and rhythm perception ($\beta = .33$, $p < .01$) represented as significant predictors. This accountability of Prosody input test scores with Tonal and Rhythm test scores coincided with the results of Pearson's correlation analysis. Interestingly, rhythm production did not significantly contribute to accounting for Prosody input test scores; instead, Rhythm production test scores significantly accounted for Focus input test scores, Chunking input test scores, and Prosody discrimination test scores. Rhythm perception test scores and Rhythm production test scores did not correlate significantly, according to Pearson's correlation analysis. Considering this non-significant correlation, rhythm production and rhythm perception may function differently in English prosody processing.

The correlation between Prosody discrimination test scores and both Tonal test scores ($r = .28$, $p < .05$) and Rhythm production test scores ($r = .36$, $p < .01$) can be explained whereby the Prosody discrimination test asked for both pitch perception and rhythm production of the English sentences. In the Prosody discrimination test, 12 pairs of English and another language each have unique syntactic rhythm and pitch movement, i.e., intonation; thereby, both rhythm sensitivity and tone sensitivity play important roles in guessing which is English in each pair. However, it should be noted that the trimmed version of the Prosody discrimination test score did not correlate with the Rhythm production test score (see Appendix P). As shown in the multiple regression analysis results in Table 28, sound sensitivity accounted for 24% of the variance in prosody discrimination with tone perception ($\beta$ = .34, $p < .05$), rhythm perception ($\beta$ = -.30, $p < .05$), and rhythm production processing ($\beta$ = .37, $p < .001$) as significant predictors. As shown in Table 29, rhythm production accounted for almost 11% of the variance in prosody discrimination as a significant predictor ($\beta$ = .33, $p < .05$). This accountability of prosody discrimination with several sound sensitivity variables, specifically rhythm production, may imply that discriminating between the prosodies of English and another language requires learners to activate and make use of sound sensitivity, particularly rhythm sensitivity. The Rhythm production test requires production of the rather long sound stimulus (see Appendix L), whose process also requires rhythmic memory. The stimulus for prosody discrimination is comparatively longer than the other English prosody processing tests (see Appendix A3). Rhythm memory might be one of the factors that correlates with Prosody discrimination test scores.

The correlation between Rhythm production test scores and Focus input test scores (*r*

= .38, $p < .01$), and the correlation between Rhythm production test scores and Chunking input test scores ($r = .36$, $p < .01$) can be explained whereby these three test scores depended partially on rhythm sensitivity. As shown in the multiple regression analysis results in Table 26, sound sensitivity accounted for 15% of the variance in chunking input, with rhythm production identified as a significant predictor ($\beta = .36$, $p < .05$). Table 27 shows that rhythm production accounted for almost 13% of the variance in chunking input as a significant predictor ($\beta = .36$, $p < .01$). As discussed in 3.6, the Chunking input test assessed the perception of prosodic boundaries of phrases created by the intensity and duration of English stress rhythm in the test phrases. From the other perspective, perception of pauses is thought to depend on knowledge of the target language (Chaippetta, Monti, & O'Connell, 1987). Chaippetta et al. (1987) reported that the nonnative speakers' duration estimates of the pauses in Italian test passages were significantly closer to the actual duration of the pauses. The nonnative speakers were significantly more accurate than the native speakers, probably because the nonnative speakers attended only to pause duration, intonation, syllable duration, level of acoustic energy, and vocalic quality as phonetic cues, whereas the native speakers relied on durational expectations using the semantic cues. Chaippetta et al. (1987) have shown that pause perception can be distracted by semantic knowledge. Thus, besides the participants' pause perception, their semantic knowledge might have influenced the scores, even under the conditions in which every participant had previous knowledge of the test phrases. In Study 2, the participants were all nonnative speakers of English and the phonetic cues might have been crucial for the pause perception of participants with less semantic knowledge, whereas the semantic cues might have been so for those with more semantic knowledge.

The correlation between Focus input test scores and Rhythm production test scores can be explained using the same deduction as in 3.6. The Focus input test asked the participants if they perceived word stresses in the English sentences. English has stress accent, and its stress pattern, i.e., its rhythm, is created with intensity, duration, and vowel quality at lexical level. The tones/pitches can be organized into postlexical phonology (Ladd, 2008). Thus, rhythm production skill could have contributed to the focus perception of English.

As shown in Table 24, sound sensitivity accounted for 16% of the variance in focus input, with rhythm production identified as a significant predictor ($\beta = .36$, $p < .01$). Table 25 shows that rhythm production accounted for 14% of the variance in focus input as a significant predictor ($\beta = .38$, $p < .01$). From the correlation analysis data, only the rhythm production skill related significantly with the focus input as a sound sensitivity variable. The Rhythm production test was intended to measure the output skill of attack points (accent), durations, dynamics, and timbres, whereas the Focus input test was intended to measure the perception of stress, i.e., prominence, in the test sentences. Therefore, it is understandable that these two test scores are significantly related.

Summarizing the results of the multiple regression analyses (Tables 22–33), it appeared that the independent variables of musical sound perception/production partially affected the dependent variables of English prosody processing skills. Although the percentage of explanations was mostly at the middle level, musical sound Perception/Production test scores, including independent variables of the Tonal test, the Rhythm test, and the Rhythm production test, significantly mostly succeeded in explaining the dependent variables of prosody input, focus input, chunking input, prosodic discrimination, and prosody production. Overall, sound sensitivity variables related with English prosody processing variables, which gives us some future perspectives as to whether and how to cultivate sound sensitivity and English prosody processing skills in elementary school education.

**4.6.2 The relationship between English prosody processing skills and English listening comprehension.** Concerning English listening comprehension as established by JET scores, JET scores and Prosody input test scores ($r$ = .36, $p$ < .01), Chunking input test scores ($r$ = .49, $p$ < .001), and Prosody production test scores ($r$ = .72, $p$ < .001) were significantly correlated. These three kinds of prosody perception/production test correlated with English listening comprehension, which might mean that prosody perception/production, i.e., intonation perception/production and pause perception, contribute to English listening comprehension.

Tables 32 and 33, and Figure 10 show the relationship between sound sensitivity and JET scores, based on collective multiple regression analysis and stepwise multiple regression analysis. Table 32 shows that English prosody processing skills accounted for almost 60% of the variance in JET scores, with prosody production as a significant predictor ($\beta$ = .64, $p$ < .001). Table 33 shows that chunking input ($\beta$ = .22, $p$ < .01) and prosody production ($\beta$ = .63, $p$ < .001) accounted for almost 56% of the variance in JET scores. Chunking input test scores and Prosody production test scores correlate significantly ($r$ = .42, $p$ < .01). This correlation can be explained by the fact that the Chunking input test assesses the perception of the prosodic boundaries in the same phrases as the material used in the Prosody production test. As stated above in 3.3 and 4.3, these tests assess the comprehension or production of the prosodic boundaries of phrases, such as “pink-and-black and green socks,” “pink and black-and-green socks,” and “chicken fingers and fruits.” Statistical correlation was inevitably found between the Chunking input test and the Prosody production test (values, see above).

As described above, pause perception may depend on knowledge of the target language

(Chaippetta et al., 1987). As the result of Study 1 (see Chapter 3), the Eiken Grade 3 listening test score and Chunking input test score significantly correlated ($r = .35$, $p < .001$). As seen in Figures 3 and 4, the pauses are made differently, in accordance with the different prosodic boundaries, in the pair of test sentences. Originally, the Chunking test intends to test the sensitivity to prosodic boundaries. The occurrence and strength of prosodic boundaries, manifested in pause duration, are systematically influenced by discourse structure, syntactic structure, phrase length, speech rate, and so on (Krivokapic, 2010). This significant result suggests that prosodic boundary perception and pause perception may rely on the semantic and syntactic knowledge required for listening comprehension, and vice versa. Thus, it may be reasonable to suggest that chunking input processing skill contributes to English listening comprehension as measured with JET.

The Prosody production test measures the reproduction skill of what listeners perceived, including pauses in English, so that achieving a high score in the Prosody production test requires English listening comprehension as a precondition. For these reasons, the prosodic boundary perception and production that are required in the Chunking input test and the Prosody production test may predict and contribute to English listening comprehension.

Structural equation modeling (SEM) was used to investigate the causal relation between sound sensitivity and English prosody processing skill, and English prosody processing and English listening comprehension. The SEM analysis converged to a solution with a good fit to the model. According to the fit indices for the model, the reasonable fit was not reflected in the model of explaining the relationship between the variables ($n$ = 53, $\chi^2(26)$ = 70.93, $p$= .000; NFI = .54; CFI = .62; RMSEA = .18). The two variables of sound sensitivity presented high factor loadings with $\beta$ = 0.80 ($p$ < .05) for the tone and $\beta$ = 0.56 ($p$ < .001) for the rhythm. The five observed variables of English prosody processing presented various factor loadings from the lowest $\beta$ = 0.22 (*ns*.) for discrimination to the highest $\beta$ = 0.85 ($p$ < .001) for prosody production. Causal relationships loading from sound sensitivity showed a significant effect on English prosody processing ($\beta$ = 0.43, $p$ < .05). Likewise, English prosody processing significantly contributed to English listening comprehension ($\beta$ = .83, $p$ < .001). Thus, English prosody processing directly affected English listening comprehension, and sound sensitivity affected English listening comprehension indirectly. Since significant relationships were found between the variables or latent variables, in spite of the poor fit indices, a better model should be found after collecting more data. The sample size should also be enlarged and this model should be analyzed in a future study.

According to these results from the correlation analysis and multiple regression analysis, several factors in high sound sensitivity appear to contribute to the improvement of English prosodic perception/production in elementary school students. High rhythm sensitivity can be one factor in the perception/production of English, considering that English prosody is regarded by Markus (2005) as stress-timed and its regular stress makes sentences sound rhythmic. The correlation between musical rhythm perception and native-like pronunciation was supported by Morgan (2003), where musical rhythm perception was measured with the same Gordon's (1982) musical aptitude test used in the present study. In Morgan's study, the significant correlation between rhythm perception and French monosyllabic vowel perception and production ($r = .30$, $p < .05$; $r = .34$, $p < .05$) was described. Morgan's study result also implies that phonological awareness and musical sound sensitivity are correlated. In the present study, in the Prosody production test, the participants' English prosody, i.e., intonation and rhythm, was mainly evaluated. In evaluating the participants' prosody production, their native-like English pronunciation of phonemes helped and enabled the evaluators to understand their English. Taking Morgan's study result into account, the correlation between native-like pronounced phonemes and musical aptitude will be a future study topic.

As far as English pronunciation is concerned, the typical English prosody of Japanese EFL learners can be heard in their function words, which are pronounced for a proportionally longer time (as described in Chapter 2). With the measurable prosodic differences between the English spoken by Japanese learners of English and native English speakers, it seems necessary for nonnative speakers of English to become conscious of their typical “Japanese-English” prosody, and improve it to rhythmic stress-timed English prosody.

The experimental studies with elementary school students provided some positive responses to the research questions. With regard to elementary school students, the following results were observed: For the first research question, musical sound perception/production does contribute to English prosodic perception and production, to some extent; for the second research question, regarding the data from the elementary school participants, musical sound sensitivity does not significantly contribute to English listening comprehension; for the third research question, English prosody perception and production contribute to English listening comprehension in second language acquisition.

Since the results from experimental Studies 1 and 2 indicate the importance of sound sensitivity in EFL from the results that sound sensitivity may contribute to English prosody processing skills, and that English prosody processing skills affect English listening comprehension, the effectiveness of early music training for enhancing sound sensitivity will be investigated in Chapter 6. In the following Chapter 5, all the data collected from the Intermediate

Measures of Music Audiation (Gordon, 1982) and PEPS-C (Peppé & McCann, 2003), as well as the Prosody discrimination test from the 53 elementary and 75 high school participants will be comparatively analyzed, with a *t*-test.

# Chapter 5 Study 3: Comparison between the Data of the High School Participants and Elementary School Participants on the Relationship between Sound Sensitivity, English Prosodic Perception, and English Listening Comprehension

## 5.1 Purpose of Study 3

Study 3 was conducted to investigate if (1) there is any difference between the sound sensitivity of the high school participants and elementary school participants, and (2) there is any difference between the English prosody perception of the high school participants and elementary school participants. If any identification or difference is found in either sound sensitivity or English prosodic perception, its cause will be discussed in the discussion section.

## 5.2 Participants

The participants comprised the 53 elementary school students and 75 public high school students who had also participated in Studies 1 and 2. The elementary school participants were 8–10 years old, and comprised 27 males and 26 females. The high school participants were 15–17 years old, and comprised 16 males and 59 females attending a public vocational high school, where students major in either agriculture or domestic science.

## 5.3 Materials

The following four materials were used for experimental Study 3 with 53 elementary and 75 high school students.

1. PEPS-C (Peppé et al., 2007): The content of this test was described in 3.3, and Appendices A1 and A2.

2. Prosody discrimination test: The content of this test was described in 3.3, and Appendices A3, A4, and E.

3. Measurement of Musical Aptitude: The content of this test was described in 3.3, and Appendices C and D.

4. Questionnaires: The questionnaires were described in 3.3 and shown in Appendix F for the high school participants, and described in 4.3 and shown in Appendices N and O for the elementary school participants.

## 5.4 Procedures

The research procedure has already been described in Chapters 3 and 4 above; however, the procedures of the elementary school participants and high school participants are combined and listed in the following order from 1 to 3.

1. On March 5, 2014 for the elementary school participants, and in the fourth week of March 2014 for the high school participants, the two major elements of musical aptitude, i.e., tonal and rhythmic processing skills, were measured by administering the Intermediate Measures of Music Audiation test in their homeroom classes in one day for each class (Gordon, 1982).

2. On June 4, 2014 for the elementary school participants, and in the fourth week of March of 2014 for the high school participants, prosodic abilities were measured with the Profiling Elements of Prosodic Systems–Children (Peppé & McCann, 2003), and English prosody discrimination skill was measured with the Prosody discrimination test, in their homeroom classes.

3. On July 4, 2014 for the elementary school participants, and in the fourth week of March 2014 for the high school participants, a questionnaire was administered on music learning background and English learning background.

### 5.5 Results

The statistical data on comparison between the elementary school participants and high school participants will be shown and brief explanations will be given on: (1) descriptive statistics, and (2) an independent *t*-test. An independent *t*-test was performed to compare sound sensitivity and English prosodic perception between the elementary and high school students. The results are shown in Table 34 below.

Before presenting the results of these statistics, uncontrolled factors such as the age and gender difference within each age group were checked with *t*-tests for the sound sensitivity tests. The result of the independent *t*-test indicated that there were no gender differences either within the elementary school participants or within the high school participants. See Appendices G, H, S, and T.

Table 34
*Independent t-Analysis on Sound Sensitivity Test Scores and Prosodic Perception Test Scores between Elementary and High School Students in Study 3*

| | Age | *N* | *M* | *SD* | *d* | *t(p)* |
|---|---|---|---|---|---|---|
| Prosody input | ES | *53* | 14.20 | 2.28 | -.03 | -.17(.87) |
| | HS | *75* | 14.28 | 3.19 | | |
| Focus input | ES | *53* | 14.96 | 1.17 | .09 | .57(.56) |
| | HS | *75* | 14.78 | 2.27 | | |
| Chunking input | ES | *53* | 11.11 | 2.61 | -.48 | 2.72**(.008) |
| | HS | *75* | 12.43 | 2.84 | | |
| Prosody discrimination | ES | *53* | 6.38 | 2.19 | -.86 | 4.78***(.00) |
| | HS | *75* | 8.27 | 2.22 | | |
| Tone | ES | *53* | 35.04 | 3.72 | .37 | 2.31*(.023) |
| | HS | *75* | 32.47 | 8.54 | | |
| Rhythm | ES | *53* | 31.60 | 4.25 | -.02 | .14(.89) |
| | HS | *75* | 31.75 | 7.50 | | |

*Note:* *$p < .05$, **$p < .01$, ***$p < .001$; *N* = the number of participants, *M* = mean score, *SD* = standard deviation, *d* = Cohen's *d*, *t* = *t*-score, *p* = probability, ES = elementary school students, HS = high school students.

## 5.6 Discussion

As shown in Table 34, there was no significant difference between the average Prosody input test scores, Focus input test scores, and Rhythm test scores of elementary and high school participants. On the other hand, significant differences were found between these two groups' Chunking input test scores ($t$ (126) = 2.72, $p$ = .008), Prosody discrimination test scores ($t$ (126) = 4.78, $p$ = .000), and Tonal test scores ($t$ (126) = 2.31, $p$ = .02). The high school participants obtained significantly higher scores in the Chunking input test and Prosody discrimination test than elementary school participants. The elementary school participants obtained significantly higher scores in the Tonal test than the high school participants.

The cause for this significant inferiority of the elementary school participants in the Chunking input test might be the uncertain prosodic boundary perception partly due to their lack of semantic and syntactic knowledge. This subtest assesses the comprehension of the prosodic boundaries of phrases, such as "pink-and-black and green socks," "pink and black-and-green socks," and "chicken fingers and fruits." Shown a pair of pictures and asked to listen to a test phrase, the participants are required to choose the picture that illustrates the pronounced phrase. Although the vocabulary composing a phrase was introduced before the test, some of the elementary school participants seemed puzzled regarding how to differentiate the pairs of

English phrases containing the same words. Some percentage of both age groups seemed to have difficulty perceiving the prosodic boundaries, lacking semantic knowledge or prosodic perception of the test phrases. Comparatively, the high school participants may have become more familiarized with the various English phrases through their junior high school and high school education. In their daily lives, the high school participants may have been more familiar with the other foreign languages such as Chinese, Korean, and French, whose prosodies were used in the Prosody discrimination test.

With regard to the result of the *t*-test on Prosody discrimination test scores of the high school participants and elementary school participants ($t(126) = 4.78$, $p < .001$), the high school participants scored significantly better. This test was intended to measure the ability to discriminate between the prosody of American/Australian English and nine other languages, including Mandarin Chinese, Russian, French, Arabic, Polish, Korean, German, Mongolian, and Japanese. In the process of listening to and discriminating the prosody pairs of the different languages, the participants might come to observe the characteristic rhythmical English prosody and become strategic in their discrimination of the following pairs. Considering the statistically same rhythm perception scores between those two groups, there should be little difference in the degree of rhythmic understanding of the test sentences. In future research, further investigation should be conducted into how much elementary school participants discriminate between the language rhythm pairs. Although they perceived the rhythm difference between pairs of test sentences, the elementary school participants might not have discriminated which was English. The high school participants might have become accustomed to English prosody through their English learning history of more than three years.

Concerning the *t*-test result of Tonal test scores, comparing the elementary school and high school participants ($t(108) = 2.31$, $p < .05$), the elementary school participants did significantly better. As Gordon (1982) proposed, musical aptitude can be enhanced until age nine, and then stabilizes, which means “a child can be expected to reach in achievement a level no higher than [a level] at which his/her potential to achieve has stabilized.” In accordance with Gordon’s proposal, the musical aptitude of both the elementary school participants and the high school participants has been stabilized around their ninth birthday. The significantly higher Tonal test scores of the elementary school participants imply that their tonal sensitivity has begun to stabilize at a significantly higher level than that of the high school participants. Out of the 53 elementary school participants, 13 were musically trained and 40 were untrained, whereas out of the 75 high school participants, only 3 were musically trained and 72 were untrained. Considering the different percentages of musically trained participants in the two groups, the significant difference in the Tonal test scores may be partly due to the participants' formal/informal music training or musical environment before age nine. In this study, the 8–10 year old elementary school participants turned out to be significantly more sensitive to the test tones than 15–17 year old participants.

Non-significant results were derived from the *t*-test on the Prosody input test ($t(126) = -.14$, *ns*), the Focus input test ($t(126) = .52$, *ns*), and the Rhythm test ($t(126) = -.13$, *ns*). This non-significance might indicate that prosody input, focus input, and rhythm can be perceived and processed to almost the same degree by both 8–10 year old elementary school students and 15–17 year old high school students. Early English education has been announced by the government and is already practiced in EFL institutions. Under such circumstances, these study results may inform the advantage of an earlier start to English prosody education, by providing an empirical example of the achievement in English focus input and prosody input processing by elementary school students.

# Chapter 6 Study 4: Effectiveness of Early Music Lessons on Enhancing Sound Sensitivity

## 6.1 Purpose of Study 4

Study 4 was conducted to answer the fourth research question (see 2.10): Do early music lessons (between the ages of three and nine, with daily music training for more than 30 minutes) affect sound sensitivity? Age nine is considered the critical cut-off point for acquiring musical aptitude, i.e., sound sensitivity (Gordon, 1982).

## 6.2 Participants

Thirty-two musically trained children and 42 untrained children aged from 6-12 years old participated in the study. The musically trained children were defined as those who had been practicing musical instruments for more than half an hour a day for more than three years. They comprised 1 guitar student, 1 flute student, 3 electric piano students, 13 piano students, and 14 violin students.

Thirty-two musically trained and 42 untrained children (74 in total) cooperated with the study. Of the 53 elementary school students who had participated in Study 2, 13 were musically trained and 40 were untrained. These 53 students were also included as participants in Study 4.

All 74 children have similar socioeconomic backgrounds. Detailed information on the participants for Study 4 is provided in Table 35.

Table 35

*Participants in Study 4*

| | Musically trained | Musically untrained | Total |
|---|---|---|---|
| Elementary school participants for Study 1, 4 | *n* = 13 | *n* = 40 | *n* = 53 |
| Elementary school participants for Study 4 | *n* = 19 | *n* = 2 | *n* = 21 |
| Total | *n* = 32 | *n* = 42 | *n* = 74 |

*Note: n* indicates the number of participants.

### 6.3 Materials

1. Musical aptitude test: The content of this test was described in 3.3, and Appendices C and D.

2. Questionnaire

This largely comprised the same questionnaire administered in Study 2 to 53 elementary school students (see 4.3). Some parts were arranged for the participants' mothers to answer. See Appendix U for the script.

## 6.4 Procedures

1. On March 5, 2014, the two major elements of musical aptitude (i.e., tone and rhythm sensitivity) were measured by administering the Intermediate Measures of Music Audiation test (Gordon, 1982) to 53 students in their two homeroom classes.

2. On July 4, 2014, the students completed the questionnaire on their music experience in their two homeroom classes (see 4.4).

3. From March to July 2014, 19 musically trained and 2 untrained children and their mothers participated in the study. Nineteen musically trained and two untrained children took the Musical aptitude test at home with the help of their mothers. The mothers returned the completed test sheets to the examiner. The mothers also cooperated in answering the questionnaires, because some of these trained children were not capable of filling them out alone.

## 6.5 Results

Tables 36 and 37 show the results of the *t*-tests that were conducted to compare tone and rhythm scores between the 32 musically trained and the 42 untrained children. Tables 36 and 37 also show the mean test scores, standard deviation, and *t*-scores. Figure 11 shows the respective densities of the Tonal test scores of the musically trained group and the untrained group. The densities were estimated with the Kernel density estimator (the Kernel density estimator indicates the estimated probability density function of the random variable). Figure 12 shows the respective densities of the Rhythm test scores of the musically trained group and the untrained group. Figures 11 and 12 reveal that there are the musically untrained participants who achieved perfect scores or nearly perfect scores, in spite of their having been untrained.

Table 36

*Independent t-Test on Tonal Test Scores of Musically Untrained and Musically Trained Children in Study 4*

| | Untrained | Musically trained |
|---|---|---|
| *N* | 42 | 32 |
| *M* | 34.17 | 38.00 |
| *SD* | 3.78 | 1.80 |
| *d* | 1.3 | |
| *t*(61.63) | 6.03*** | |

*Note:* Equal variances were not assumed, *d* indicates effect size, i.e., Cohen's *d*; ****p* < .001.

Table 37

*Independent t-Test on Rhythm Scores of Musically Untrained and Musically Trained Children in Study 4*

| | Untrained | Musically trained |
|---|---|---|
| *N* | 42 | 32 |
| *M* | 30.93 | 34.19 |
| *SD* | 4.41 | 2.89 |
| *d* | 0.89 | |
| *t*(64) | 4.12*** | |

*Note:* Equal variances were not assumed; *d* indicates effect size, i.e., Cohen's *d*; ****p* < .001.

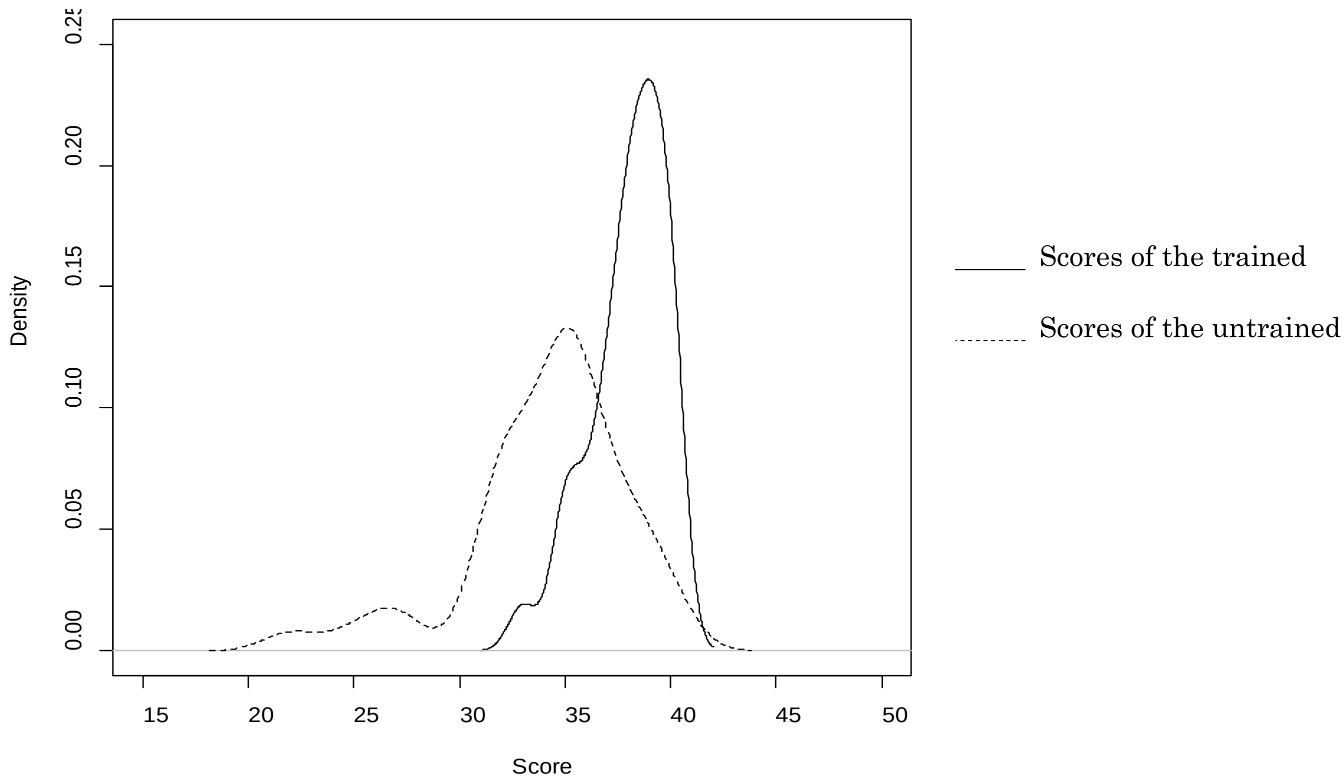



*Figure 11.* The density of Tonal test scores of two groups in Study 4.

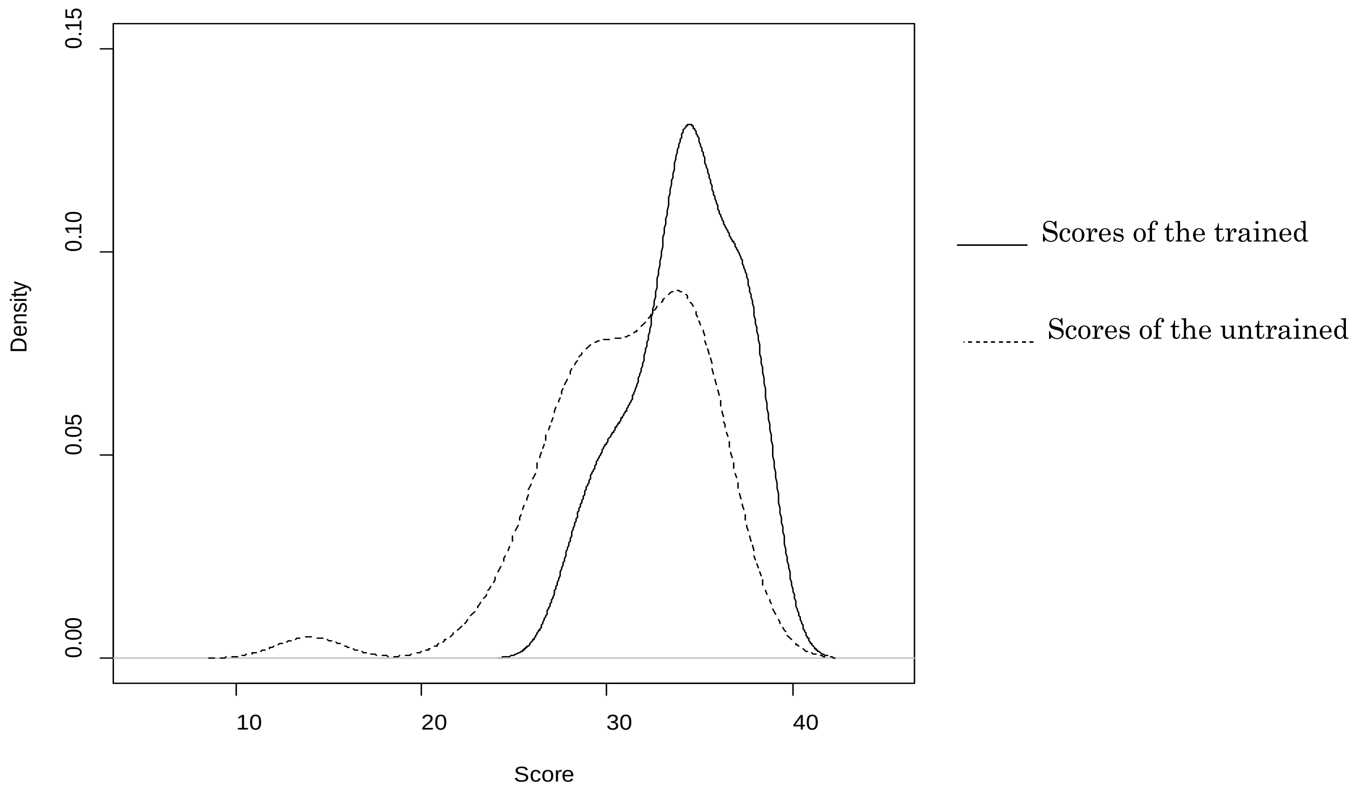



*Figure 12.* The density of Rhythm test scores of two groups in Study 4.

## 6.6 Discussion

According to the results of the *t*-tests shown in Tables 36 and 37, there was a significant difference in sound sensitivity, represented by the Tonal and Rhythm test scores, between musically untrained and musically trained children. The significant differences in musical aptitude were calculated as $t$ (61.63) = 6.03, $p < .001$ in Tonal test scores, and $t$ (64) = 4.12, $p < .001$ in Rhythm test scores, implying that the musically trained children were significantly better at tone/rhythm perception. In particular, regarding the Tonal test scores, their musical experience seems to have had a strong influence on the participants, with a large effect size

(Cohen's *d* = 1.3). However, there are the untrained participants who had perfect or nearly perfect scores; thus, it can be assumed that music training needs to be provided to those who are both untrained and obtained relatively low scores in the musical aptitude tests. As previous researchers have demonstrated the effectiveness of early music training (Gordon, 1982; Lynch & Eilers, 1991; Magne et al., 2006), the results of Study 4 provided a positive answer to the research question: Do early music lessons (for three or more years before age nine, with daily music training for longer than 30 minutes) affect sound sensitivity? The answer is "Yes."

Previous researchers have debated the critical and sensitive periods for music acquisition (Gordon, 1982; Trainor, 2005). The critical period hypothesis and sensitive period (Long, 2013a) have been discussed with regard to their applicability to sound sensitivity acquisition, particularly to absolute pitch acquisition. Crozier (1997) suggested that absolute pitch is acquired genetically or before the age of six. This suggestion supports the effectiveness of early music training in enhancing sound sensitivity.

In Japan, music education, including instrumental music, starts from the first year of elementary school (MEXT, 2015b). Musical instruments practiced in the first and second year are defined as "familiar musical instruments to pupils," such as melodicas, castanets, tambourines, triangles, and various kinds of drums, including traditional Japanese drums (MEXT, 2015b). Considering the potential benefits of early music training (Schellenburg, 2003), it seems worthwhile beginning music education formally or informally in early childhood. Gordon (1982) stated the importance of quality in formal music instruction in the classroom, particularly up to the third year of elementary school, and of continuing interaction with informal environmental influences. Gordon (1982) also emphasized the need to provide informal environmental influences in preparation for formal music instruction. Thus, in addition to providing fruitful opportunities to listen and perform music from the beginning of compulsory education, or even in preschool education, an ideal quality of the formal and informal music environment needs to be aimed at both inside and outside educational institutions.

For enhancing sound sensitivity in EFL, elementary music education should emphasize on Western instrumental music. The effectiveness of early Western instrumental music training for acquiring better musical aptitude was empirically implied in Study 4, particularly in the low aptitude population. The relationships between sound sensitivity, English prosody processing, and English comprehension investigated through this thesis proved that sound sensitivity contributes to English prosody processing skill and indirectly contributes to English listening comprehension. In a musically enriched environment, sound sensitivity is sure to be enhanced and to contribute to prosody processing in EFL.

# Chapter 7 Conclusion and Implications

## 7.1 Conclusion and Implications for Japanese EFL Education

According to the results and discussion in Studies 1–4, research questions (1)–(4) were partly answered in the affirmative. The affirmations provided by this research can be summarized as follows: (1) Higher sound sensitivity may contribute to better English prosody processing in 8–10 year old elementary school students and 15–17 year old high school students; (2) Higher sound sensitivity might contribute implicitly to better English listening comprehension in 15–17 year old high school students (see 3.5.4, Figure 9); (3) Better English prosody processing might contribute to better English listening comprehension in second language acquisition in 8–10 year old elementary school students and 15–17 year old high school students; (4) Early music lessons (between the ages of three and nine, with daily music training for more than 30 minutes) might affect sound sensitivity.

**7.1.1 In relation to the SEM model of the high school students.** Overviewing the findings from the correlation analysis, multiple regression analysis, and SEM analysis in Study 1, all reached the conclusion that there was some relation between sound sensitivity, English prosody processing skill, and English listening comprehension for the 15–17 year old high school participants. Above all, the SEM explains this relation visually.

Simply reviewing the results and implications related to the SEM model (see Figure 9 in 3.5.4), the first finding is that sound sensitivity can relate to English prosody perception: From the stepwise multiple regression analyses, rhythm perception had a significant effect on the accountability for prosody discrimination ($\beta$ = .38, $p$ < .01), focus input ($\beta$ = 47, $p$ < .001), chunking input ($\beta$ = .26, $p$ < .05), and Eiken Grade 3 listening test scores ($\beta$ = .31, $p$ < .01). This may suggest the direct contribution of rhythm perception to English listening comprehension. Tone perception explained prosody input ($\beta$ = .38, $p$ < .01). Considering these accountabilities of rhythm perception for English prosodic processing skill, the enhancement of high rhythm sensitivity can contribute to better perception of English rhythmical prosody.

Comparing the rhythmic-prosodic structure of speech between English and Japanese, it is recalled that English is a stress-timed language, whose regularly arising stresses make sentences sound rhythmic. In contrast, Japanese is generally understood as a mora-timed language, a variation of a syllable-timed language (Satoi, 2003; Grabe, 2002). The rhythmic-prosodic structural differences between English and Japanese may cause the difficulties in perception and production of English that are often experienced by elder Japanese learners. In terms of prosodic character, mora-timed Japanese has quite a simple syllable structure and no vowel reduction;

thus, Japanese EFL learners with poor metrical prosody perception should be unable to comprehend English easily. The positive transfer from sound processing initially to English prosody processing, and English prosody processing eventually to English listening comprehension ought to be realized in Japanese EFL learners' learning process.

The second finding from Study 1 has implied the existence of a mid-level correlation between English prosody perception skills and Eiken Grade 3 listening test scores. However, the three factors (i.e., prosody input, focus input, and chunking input) that were correlated significantly with listening comprehension were also all highly significantly correlated with one another. Because of this, it may be difficult to adequately discuss the independent contribution of each individual factor. Despite this difficulty in terms of making definite statements, the statistical results will be useful to grasp some tendency.

Among the significant Pearson's correlation analysis results, the strongest correlation was found between Eiken Grade 3 listening scores and focus input ($r = .50$, $p < .001$). Given this data, the perception of focus in a sentence should be considered an important prosodic cue for sentence comprehension in stress-timed languages and it may have a rather strong relation with English listening comprehension. Improvement of learners' perception of nuclear stress in phrases or sentences may result in better English listening comprehension.

The second strongest correlation was found between Eiken Grade 3 listening test scores and prosody input ($r$ = .40, $p$ < .001). This implies that the perception of phrasal or sentence intonation, that is, intonational prosody, may help learners predict its syntactic and semantic information. As mentioned earlier in this paper, Ichikawa (2011) describes how speech prosody adds information at the time the speaker makes an utterance. In this light, Ichikawa (2011) proposes the existence of “message preliminary announcement information” as understood via speech prosody. Ichikawa’s analysis and proposal support the implication found in the present study that prosodic perception contributes to the listening comprehension of English.

The third largest correlation was observed between the Eiken Grade 3 listening test score and chunking perception ($r$ = .35, $p$ < .001). Chunking, as reflected in pauses, is considered by Goldman-Eisler (1972) to express prosodic cues. Similarly, Butcher (1981) states that the grammatical functions of prosody include the segmenting of utterances into prosodic phrases; the ends of phrases are signaled by a number of prosodic factors, including pausing after the phrase. Other researchers have defined prosodic cues for segmentation as “the combination of pitch change and pre-boundary lengthening, or the combination of pitch change and pause” (Scott, 1982), or as “the inclusion of the nuclear tone at or near the end of the utterance” (Crystal, 1969). Thus, segmentation and chunking are different concepts expressing similar phenomena, and should be viewed as sources of prosodic cues to help listening comprehension in English.

Of course, these positive relationships between sound sensitivity, English prosody perception, and English listening comprehension could potentially be explained by other plausible factors, such as overall academic performance, test-taking strategy, motivation, and so on. It is a limitation of the present study that those potential factors could not be factored out as covariates.

Above all, the most prominent finding from Study 1 was that the hypothesized relation between latent variables–sound sensitivity, English prosody processing variables, and English listening comprehension–were statistically asserted with the SEM model (see Figure 9 in 3.5.4). This finding sheds lights on the possibility of establishing a link between music education and foreign language education in enhancing the sound sensitivity for both music and language.

**7.1.2 In relation to the multiple regression model of the elementary school students.** As the implications of Study 2, firstly, sound sensitivity, including tone sensitivity and rhythm sensitivity, may relate to English prosody processing skill in 8–10 year old elementary school students. Secondly, English prosody processing skill may relate to English listening comprehension in 8–10 year old elementary school students. The collective multiple regression analysis and stepwise multiple regression analysis were conducted with sound sensitivity, comprising tone perception, rhythm perception, and rhythm production as predictors of English prosody processing skills, consisting of prosody input, focus input, chunking input, prosody discrimination, and prosody production.

As a result, sound sensitivity accounted for almost 38% of the variance in prosody input, with tone perception ($\beta = .39$, $p < .01$) and rhythm perception ($\beta = .32$, $p < .05$) being identified as significant predictors (see Table 22). As a similar positive result, sound sensitivity accounted for 38% of the variance in prosody input with tone perception ($\beta = .39$, $p < .001$) and rhythm perception ($\beta = .33$, $p < .01$) represented as significant predictors (see Table 23). As described in 4.6, the accountability of Prosody input test scores with Tonal and Rhythm test scores coincided with the significant correlations among these three variables (see Table 21 for Pearson's correlation analysis). It also partially coincided with the result of stepwise multiple regression analysis on the high school participants, whose Prosody input test scores were significantly accounted for by Tonal test scores ($R^2 = .14$, $\beta = .38$, $p < .01$).

As shown in Table 28, sound sensitivity accounted for 24% of the variance in prosody discrimination with tone perception ($\beta = .34$, $p < .05$), rhythm perception ($\beta = -.30$, $p < .05$), and productive rhythm processing ($\beta = .37$, $p < .001$) as significant predictors. Although rhythm perception was not positively related with prosody discrimination, which belied expectations, the other sound sensitivity variables, particularly rhythm production, were related positively with prosody discrimination. Thus, sound sensitivity may contribute to English prosody processing in some way.

Concerning the relation between the English prosody processing skills and English listening comprehension of the elementary school students, the collective multiple regression analyses show that English prosody processing skill accounted for almost 60% of the variance in JET scores, with prosody production as a significant predictor ($\beta = .64$, $p < .001$). As described in 4.3, reproduction of spoken phrases and pausing appropriately between the phrases was tested in the Prosody production test. Butterworth (1975) analyzed the distributions of phonations and pauses in spontaneous speech, and found a rhythmic cycle that reflected the cognitive process of the speaker. In order to reproduce a speaker's rhythmic cycles in speech, the prosody perception of speech may be a prerequisite and prosody production skill should be requisite. According to the stepwise multiple regression analysis shown in Table 33, chunking input ($\beta = .22$, $p < .01$) and prosody production processing ($\beta = .63$, $p < .001$) accounted for almost 56% of the variance in JET score as significant predictors. This result also implies that English listening comprehension may presuppose prosody processing skills such as chunking input and prosody production.

Although the hypothesized SEM model did not have a good fit to the data of elementary school participants, the multiple regression model (see Figure 10) revealed the contribution of sound sensitivity variables to English prosody processing variables, and the other contributions of English prosody processing variables to English listening comprehension.

**7.1.3 In relation to the comparative model between the high school students and the elementary school students.** As an implication of Study 3, sound sensitivity may be acquired until age nine and does not drastically change after that, referring to the result that there was no significant difference in rhythm sensitivity between the 8–9 year old participants and the 15–16 year old participants. The other result showed that tone sensitivity was significantly better in the 8–9 year old participants than in the 15–16 year old participants. Simply observing this result of better tone perception in the younger participants, the 8–9 year old elementary school participants might possess better tone sensitivity than the 15–16 year old high school participants. However, considering that the number of musically trained children among the 8–9 year old elementary school participants ($N$ = 13) is larger than that for the 15–16 year old high school participants ($N$ = 3), the experimental study comparing the 53 and 75 participants may not be sufficient to generalize the result, and we must await a future longitudinal study with more participants.

The other implication from Study 3 is that chunking input and prosody discrimination were perceived better by high school participants than elementary school participants. As described in 5.6, the prosodic boundary perception and pause perception required in the Chunking input test and the Prosody discrimination test may require a certain amount of exposure to the target language. As far as the previous research on pausing is concerned, it has been studied in its overall proportions in speech. It was found that the distribution of pause durations at sentence boundaries, in both spontaneous speech and reading, was quite characteristic, and unlike distributions for other syntactic transition points (Goldman-Eisler, 1972). Goldman-Eisler (1972) defined the rhythmic patterns of speech where the amount of speech was dependent mathematically on the amount of pausing. According to Goldman-Eisler (1972) and Butterworth (1975), the amount of pausing in overall proportions in speech, in its durations and locations, is estimated mathematically, cognitively, and psychologically during spontaneous speech. In this respect, pausing duration and location in a sentence or phrase may be estimated relative to the whole sentence or phrase, concerning the overall proportion of pausing in them. When perceiving pausing in spontaneous speech, therefore, experienced listeners may become more accustomed to English rhythmic patterns comprised of phonations and pauses. In this study, the elementary school participants may not have been accustomed to English rhythmic patterns for pause perception nor have obtained the semantic and syntactic knowledge required for achieving a high score in the Chunking test score. It is assumed that perception of prosodic boundaries and pauses measured in the Chunking input test was significantly better among the high school participants, due to their larger amount of experience and knowledge.

The Prosody discrimination test was intended to measure the ability to perceive the rhythmic patterns of the given English sentences. The significantly higher scores of the high school participants may imply that experienced learners have become familiar with the rhythmic pattern of English with voiced phonations and silent pauses, and with characteristic prosodies of broadly-penetrated languages such as Chinese, Korean, and French. On the other hand, the elementary school participants may not have been sufficiently familiar with English rhythm or the rhythms of the other foreign languages to recognize the rhythmic patterns by their prosody alone, even if they observed the rhythm difference in the prosody sentence pairs.

**7.1.4 In relation to the early music training effect on sound sensitivity.** Study 4 implied that early music training can implicitly and indirectly contribute to the sound sensitivity of learners under nine years old. As precisely described in Chapter 6, the hypothesis was supported that early musical lessons (between the ages of three and nine, with daily music training for 30 minutes at a minimum) affect sound sensitivity. The *t*-test results demonstrated a significant difference in sound sensitivity, both in tone scores ($t$ (62) = 6.03, $p < .001$) and rhythm scores ($t$ (64) = 4.12, $p < .001$) between the trained and untrained participants.

From another perspective, referring to Figures 11 and 12, some of the musically untrained children scored perfectly or nearly perfectly, which indicates that musical training will work meaningfully not for children with already high musical aptitude but for children with low musical aptitude. The asserted hypothesis will be discussed in respect of the effectiveness of early sound sensitivity training, in the future study section 7.3.

## 7.2 Limitations of the Study

One of the limitations common to Studies 1–4 is the sample size. In an empirical research, a large sample should be collected and analyzed statistically, in order to obtain reliable and valid data and results. However, in general elementary schools and high schools in Japan, pupils cannot devote a large amount of time to educational research. The participants and their schools in this research offered a few hours for the author’s educational purpose. Their cooperation made this research successful. Additionally, a longitudinal study over a broader age range, from infants to middle-aged adults, will be more appropriate in order to observe how musical aptitude, English prosody processing, and English listening comprehension improve in controlled and uncontrolled situations.

The other limitation concerned the JET, which consists of 25 listening test questions and 5 vocabulary questions. In total, the 30 questions were given to the participants, and each total score was evaluated as representing listening comprehension skill in this study. However, the actual test score is 100 points in total, calculated with a scoring system that is owned by JET. Thus, the JET score in this study of approximately 16% was related to participants' vocabulary knowledge. In this study, reading question scores are included in listening comprehension scores; thus, they are not genuine listening comprehension scores. The total JET scores were provided as data, and 87% of each score precisely reflected the participants' listening comprehension of English. Therefore, the author considered and used this as the English listening comprehension test. In a future study, listening comprehension scores should be carefully extracted as data.

This dissertation has concentrated mainly on confirming the relationships between sound sensitivity, English prosody perception and production, English listening comprehension, and early musical training. The implications stemming from these relationships point to the need for earlier childhood language and music instruction, resulting in better overall learning and learners. While my references and student data from the many tests empirically support this understanding, there is also a more diverse body of research (that is beyond the scope of this thesis) that would also support these findings in 1) infant development (Krueger, 2013), 2) anthropology (Mithen, 2006), 3) medicine (Sacks, 2007), and 4) neuroscience (Patel, 2008; Kraus & Slater, 2016).

## 7.3 Suggestions for Future Research

In this section, suggestions will be made for future studies, initially in the following two respects: Firstly, a suggestion for studies of overall English proficiency; secondly, a suggestion for determining the best period for starting music and EFL education. Then, three follow-up suggestions will be made to round up this dissertation. These suggestions are: a) to examine the score differences among the 12 questions in the Prosody discrimination test with a variety of foreign languages' prosodies and a variety of participant ages; b) to examine why the Focus input test results contributed to listening comprehension by different degrees between the respective multiple regression models of high school and elementary school participants; and c) to more closely examine other possible reasons for the overall statistical results of Studies 1-4, such as reasons relating to the wealth and health of participants.

**7.3.1 A suggestion for studies on overall English proficiency.** Although further evidence and confirmation are required, the affirmative results above imply that empowerment in sound processing skill will result in better English prosody processing and better EFL listening comprehension in 8–10 year old elementary school students and 15–17 year old high school students. Prospecting an earlier EFL start in Japanese elementary schools, in order for elementary school aged children to acquire communicative English skills, further research should be conducted on the relationship between musical sound sensitivity, English prosodic processing skills, and practical English communication skills such as listening and speaking proficiency. In this research, as regards total proficiency in English, listening comprehension was mainly tested in terms of its correlational relationship with sound perception/production and English prosody perception/production. The other skills, such as reading, writing, and speaking, should also be examined as dependent variables. The relationships between sound perception/production, English prosody perception/production, and inclusive communicative English competence, including listening, speaking, reading, and writing skills, would be a suitable topic for further research.

**7.3.2 A suggestion for determining the best period for starting music and EFL education.** Further research must be conducted on how to cultivate English prosody processing skill and musical sensitivity through Japanese compulsory education classes. In the current guidelines announced by MEXT (2015a, 2015b), music education starts from the first year of elementary school, and comparatively, foreign language activity starts from the fifth year of elementary school. The overall objective of foreign language activities is described as “to form the foundation of pupils’ abilities through foreign language while developing the understanding of language and cultures through various experiences, fostering a positive attitude toward communication, and familiarizing pupils with the sounds and basic experiences of foreign language” (MEXT, 2015a, p. 1). The phrase in the previous sentence “familiarizing pupils with the sounds of foreign language” is equivalent to what has been discussed through this thesis. Although further evidence is required, the results from Studies 2 and 4 indicate that sound sensitivity relates with English prosody processing skill in 8–10 year old children, and that early music training can enhance sound sensitivity and implicitly improve their English prosody processing skill. To help children acquire communicative English skills, further research should investigate the practical methodology that transfers the effect of musical education to language prosody education and ultimately to communicative language education.

In addition, research is required to investigate when to cultivate English prosodic ability and musical aptitude in Japanese elementary school classes. A statistical correlation was found between musical aptitude and English perceptive/productive prosody skills in 8–10 year old children, which indicates a probable crucial/sensitive period for these skills before or within this range. More longitudinal and cross-sectional research should be conducted in the future to identify the most appropriate time to start training in these skills.

Previous researchers have debated the critical period hypothesis for music acquisition (Trainor, 2005) and language acquisition (Krashen, Long, & Scarcella, 1979). DeKeyser (2000) implies that the critical age for language learning is as follows: “The decline of language learning ability seems to take place gradually from ages 6–7 to 16–17 and beyond” (p. 500). As stated above, in Japan, public English education usually begins in fifth grade, and music education in first grade (MEXT, 2015b). Given the potential benefits of early English learning and music training (Schellenburg, 2003), linguistic and musical prosody education seems worth starting in early childhood, although more research should be conducted on the mechanism and methodology for early linguistic and musical prosody acquisition.

**7.3.3 A suggestion for future study.** Firstly, further statistical analysis will be necessary to look at the score differences between the 12 questions with a variety of foreign language prosodies in the Prosody discrimination test, and the reason for the significantly-better scores of high school participants than those of elementary school participants in this. According to the regression tree analysis, language prosody types (stress-timed, syllable-timed, and mora-timed) were found not to be significant variables with regard to the high school participants’ scores on the Prosody discrimination test (see Appendix V). Further study is necessary to conduct a comparison between the nine language pairs to find out whether a familiarity with certain foreign languages or their rhythm characteristics contributed to the results of high school participants' performance on the Prosody discrimination test.

Secondly, there may be two reasons that the Focus input test results contributed to listening comprehension by different degrees between the respective multiple regression models of high school and elementary school participants. One reason may be the different English learning histories between high school and elementary school participants. The other reason may be the different measurement materials for English listening comprehension: the Eiken Grade 3 Listening Test for the high school participants and JET for the elementary school participants. These possible reasons should be qualitatively and quantitatively analyzed in a future study.

While this dissertation finds that early childhood music and language lessons support better English processing and production, there are two other factors that I suspect might also have a significant impact on these results and they need to be controlled for in future studies; these two factors are the health and wealth of the participants. Good socio-economic conditions may favor some children by providing easy access to music and foreign language lessons in cram and recreational schools, thus instilling possible sound sensitivity, while underprivileged children may not possess the financial means to access these opportunities. At the same time, even among those children with the socio-economic means to access these lessons, some may suffer from undiagnosed hearing deficiencies or biological deficiencies (a lack of healthy hearing) that keep them from engaging effectively in learning the sound parameters of music and foreign languages. Therefore, it could be that children who continue with lessons over time are both from higher socio-economic levels and reasonably healthy, whereas those without the means (wealth and health) have less chances to acquire sound sensitivity and skills that accumulate over years of learning.

## References

Algeo, J. (1992). American English. In T. McArthur (Ed.) *The Oxford companion to the English language* (pp. 37–41). Oxford, UK: Oxford University Press.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning, 42*(4), 529–555.

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83*, 111–130.

Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition: Acoustic analysis of duration and F0 range. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 281–297). Amsterdam: John Benjamins.

Arvaniti, A., Ladd, R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics, 26*, 3–25.

Atterbury, B. W. (1985). Musical differences in learning-disabled and normal-achieving readers aged seven, eight and nine. *Psychology of Music, 13*, 114–123.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in English and Japanese.

*Phonology Yearbook*, *3*, 255–310.

Besson, M., & Faïta, F. (1995). An Event-Related Potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology and Performance*, *21*(6), 1278–1296.

Besson, M., Magne, C., Moreno, S., Schön, D., & Santos, A. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, *25*, 399–410.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9–10), 341–345.

Butcher, A. (1981). *Aspects of the speech pause: Phonetic correlates and communicative function.* Kiel: Institut für Phonetik der Universität Kiel.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *4*(1), 75–87.

Call, M. E. (1985). Auditory short-term memory, listening comprehension, and the Input Hypothesis, *TESOL Quarterly, 19*(4), 765–781.

Chaippetta, J., Monti, L., & O'Connell, D. (1987). Pause perception: Some cross-linguistic comparisons. *Psychonomic Society*, *25*(2), 103–105.

Chan, A., Ho, Y., & Cheung, M. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology, 17, 3,* 439-450.

Chang, E., & Merzenich, M. (2003). Environmental noise retards auditory cortical development. *Science, 300*, 498–502.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

Cooper, W. E., & Sorenson, J. M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of Acoustic Society of America, 62*, 683–692.

Crozier, W. R. (1997). Absolute pitch: Practice makes perfect, the earlier the better. *Psychology of Music, 25*, 110–119.

Crystal, D. (1969). *Prosodic systems and intonation in English.* Cambridge, UK: Cambridge University Press.

Darwin, C. (1872). *The expression of emotion in man and animals.* London, UK: John Murray.

Defense Language Institute Foreign Language Center. (2009). *Defense language proficiency testing system 5 framework*. Retrieved from http://www.dliflc.edu/file.ashx?pass=archive/documents/Framework_Document_Se p_10_09.pdf.

DeKeyser, M. R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499–533.

Dominguez, R., & Pessoa, S. (2005). Early versus late start in foreign language education: Documenting achievements. *Foreign Language Annals, 38*(4), 473–483.

Doughty, C. (2013). Optimizing post-critical-period language learning. In G. Granena & M. Long (Eds.), *Sensitive periods, and ultimate L2 attainment. Language Learning & Language Teaching, 35*, (pp. 231–258). Philadelphia, PA: John Benjamins.

Doughty, C., Campbell, S., Bunting, M., Mislevy, M., Bowles, A., & Koeth, J. (2010). Predicting near-native L2 ability: The factor structure and reliability of Hi-LAB. In M. T. Prior et al. (Eds.), *Proceedings of the 2008 Second Language Research Forum* (pp. 10–31). Somerville, MA: Cascadilla Proceedings Project. Retrieved from http://www.lingref.com/cpp/slrf/2008/index.html.

Douglas, S., & Willats, P. (1994). The relationship between musical ability and literacy skills. *Journal of Research in Reading, 17*, 99–107.

Dunbar, R. (1997). *Grooming, gossip and the evolution of language.* New York, NY: Faber & Faber.

Fok, C. Y-Y. (1974). *A perceptual study of tones in Cantonese.* Hong Kong: Centre of Asian Studies, University of Hong Kong.

François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex, 21*, 2357–2365.

François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hearing Research, 308*, 122–128.

Friedrich, M. (2004). The infant's brain response to words and nonsense words in meaningful contexts. Talk presented at *Architectures and Mechanisms for Language Processing* (AMLaP), 16 September 2004, Aix-en-Provence, France.

Fujisaki, H. (1996). Prosody, models, and spontaneous speech. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing Prosody* (pp. 27–42). New York, NY: Springer-Verlag.

Gilbert, J. (2008). *Teaching pronunciation: Using the prosody pyramid.* New York, NY: Cambridge University Press.

Gimson, A. C. (1962). *An introduction to the pronunciation of English*. New York, NY: St. Martin's Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, *15*, 103–113.

Gordon, E. E. (1982). *Intermediate measures of music audiation.* Chicago, IL: G.I.A. Publications.

Gordon, E. E. (1994, December–January). Audiation, the door to musical creativity. *Pastoral Music*, 39–41.

Gottfried, L. T. (2007). Music and language learning. Effect of musical training on learning L2 speech contrasts. In B. Ocke-Schwen & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 221–237). Amsterdam: John Benjamins.

Grabe, E. (2002). *Variation adds to prosodic typology.* In B. Bel & I. Marlin (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 127–132), 11–13 April 2002, Aix-en-Provence: Laboratoire Parole et Langage.

Huron, D. (2003). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences, 930,* 43–61.

Ichikawa, K. (2011). *The science of the language of dialogs* --- *Prosody facilitates communication.* Waseda University Press.

Ichikawa, U. (2007). *Rhythm training*. Chuo Art Publisher.

Inoue, F. (2006). Sociolinguistic characteristics of intonation. In Y. Kawaguchi, I. Fonagy, & T. Moriguchi (Eds.), *Prosody and syntax: Cross-linguistic perspective* (pp. 197–222).

Amsterdam: John Benjamins.

Jackendoff, R., & Lerdahl, F. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Jacobsen, M., & Imhoof, M. (1974). Predicting success in learning a second language. *Modern Language Journal*, *58*, 329–336.

Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals.* Oxford, UK: Oxford University Press.

Johnson, S. J., & Newport, L. E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics in Language Teaching*, *28*(2), 99–118.

Jun, S.-A. (2012). Prosodic typology. In: S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430-458). Oxford, UK: Oxford University.

Jusczyk, W. P., Palmer, C., & Jungers, K. M. (2001). Episodic memory for musical prosody. *Journal of Memory and Languages, 45,* 526–545.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse.

*Journal of Phonetics*, *3*, 129–140.

Koelsch, S., Gunter, T., Cramon, D., Zysset, S., Lohmann, G., & Friederici, A. (2002). Back speaks: A cortical “language-network” serves the processing of music. *NeuroImage*, *17*, 956–966.

Koine, Y. (1995). New English-Japanese dictionary. Tokyo, Japan: Kenkyusha.

Krashen, S. (1977). Some issues relating to the monitor model. In H. Brown, C. Yorio, & R. Crymes (Eds.), *Teaching and learning English as a Second Language: Trends in research and practice* (pp. 144–215). Washington, DC: TESOL.

Krashen, S. D., Long, M., & Scarcella, R. (1979). Age, rate, and eventual attainment in second language acquisition. *TESOL Quarterly, 9*, 573–582.

Krashen, S. D., Terrell, T. D., Ehrman, M. E., & Herzog, M. (1984). A theoretical basis for teaching the receptive skills. *Foreign Language Annals*, *17*(4), 261–275.

Kraus, N., & Slater J. (2016). Beyond words: How humans communicate through sound. *Annual Review of Psychology*, *67*, 83–103

Krivokapic, J. (2010). Speech planning and prosodic phrase length. *Every Language, Every Style: Speech Prosody, 5th International Conference*. Chicago, IL. Retrieved from speechprosody2010.illinois.edu/papers/100311.pdf.

Krueger, J. (2013). Empathy, enaction, and shared musical experience: Evidence from infant cognition. In T. Cochrane, B. Fantini, & K. Scherer (Eds.), *The emotional power of music: Multidisciplinary perspectives on musical expression, arousal, and social control* (pp. 177–196). Oxford: Oxford University Press.

Ladd, R. (1996). *Intonational phonology.* Cambridge, UK: Cambridge University Press.

Ladd, R., Dan, F., Hanneke, F., & Schepman, A., (1999). Constant segmental anchoring of F0 movements under changes in speech rate. *Journal of acoustic Society of America 106*, 2685-2696.

Ladd, R. (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Rivista di Linguistica, 18*(1), 19–38.

Ladd, R. (2008). *Intonational phonology.* Cambridge, UK: Cambridge University Press.

Lamb, S. J., & Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educational Psychology, 13,* 262–266.

Levintin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage, 20,* 2142–2152.

Liberman, M. (1975). *The intonational system of English* (Unpublished doctoral dissertation).

Massachusetts Institute of Technology, Cambridge, MA.

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

Linck, J., Hughes, M., Campbell, S., Silbert, N., Tare, M., Jackson, S., Smith, B., Bunting, M., & Doughty, C. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*(3), 530–566.

Long, M. (1990a). The least a second language acquisition theory needs to explain. *TESOL Quarterly, 24,* 649–665.

Long, M. (1990b). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251–285.

Long, M. (2013a). Some implications of research findings on sensitive periods in language learning for educational policy and practice. In G. Granena & M. Long (Eds.), *Sensitive periods, and ultimate L2 attainment. Language Learning & Language Teaching, 35,* 259–271. Philadelphia, PA: John Benjamins.

Low, E., & Grabe, E. (1995). Prosodic patterns in Singapore English. *Proceedings of the International Congress of Phonetic Sciences, Stockholm, Sweden, 3,* 1636–1639.

Low, E., & Grabe, E. (2002). Durational variability in speech and rhythm class hypothesis. In C.

Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515–546). Berlin: Mouton De Gruyter.

Lynch, M. P., & Eilers, R. E. (1991). Children's perception of native and non-native musical scales. *Music Perception, 9*, 121–132.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. (2001). Musical syntax is processed in Broca's area. *Nature Neuroscience, 4*(5), 540–545.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18*(2), 199–211.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Neurocognitive Science, 23*(10), 2701–2715.

Markus, M. (2005). English and German prosody – A constructive comparison. In Y. Kawaguchi, I. Fónagy, & T. Moriguchi (Eds.), *Prosody and syntax* (pp. 103–140). Amsterdam: John Benjamins.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463.

Martin, P. (1996). *WinPitch LTL, un logiciel multimédia d'enseignement de la prosodie*. Retrieved from http://alsic.revues.org/332; DOI: 10.4000/alsic.332.

Meerman, A. D., Kiyama, S., & Tamaoka, K. (2014). To what extent does accent sensitivity provide the foundation for lexical knowledge and listening comprehension? *Open Journal of Modern Linguistics*, *4*, 457–464.

Milovanov, R., Huotilainen, M., Esquef, P., Alku, P., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience Letters*, *460*, 161–165.

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

Ministry of Education, Culture, Sports, Science and Technology. (2015a). *Course of study on elementary school foreign language activities.* Retrieved from http://www.mext.go.jp/component/english/__icsFiles/afieldfile/2011/03/17/1303755_007.pdf.

Ministry of Education, Culture, Sports, Science and Technology. (2015b). *Course of study on* *elementary school of music.* Retrieved from http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2009/04/21/1261037_7.pdf

Mithen, S. (2006). *The singing Neanderthals.* Cambridge, MA: Harvard University Press.

Morgan, C. (2003). *Musical aptitude and second-language phonetics learning: Implications for teaching methodology* (Unpublished doctoral dissertation). Simon Fraser University, British Columbia.

Muñoz, C. (2010). On how age affects foreign language learning. *Advances in Research on Language Acquisition and Teaching: Selected Papers*. Retrieved from http://www.enl.auth.gr/gala/14th/Papers/Invited%20Speakers/Munoz.pdf

Nakano, H. (2012). *Eikaiwa rhythm method.* Obunsha.

Nakano, H., & Natsume, K. (2011). English rhythm learning and changes in EEG using RIM with beat. *Computers and Education, 31*, 88–93.

Neufeld, G. (1978). On the acquisition of prosodic and articulatory features in adult language learning. *The Canadian Modern Language Review, 34*, 163–174.

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

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music.

*Cognition, 87*, B35–B45.

Peppé, S., & McCann, J. (2003). *Profiling elements of prosody in speech-communication (PEPS-C).* Retrieved from http://www.peps-c.com/index.html.

Peppé, S., McCann, J., & Gibbon, F. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *50,* 1015–1028.

Pesetsky, D. (2008). *The syntax of music syntax is the syntax of language* (Unpublished doctoral dissertation). University of Utrecht, Netherlands.

Peynircioğlu, Z. F. (2002). Phonological awareness and musical aptitude. *Journal of Research in Reading, 25*(1), 68–80.

Pimsleur, P., Stockwell, R. P., & Comrey, A. L. (1962). Foreign language learning ability. *Journal of Educational Psychology, 53*(1), 15–26.

Pimsleur, P. (1966). *The Pimsleur language aptitude battery.* N. Bethesda, Maryland: Second Language Testing, Inc.

Pimsleur, P., Reed, D. J., & Stansfield, C. W. (2004). *Pimsleur Language Aptitude Battery: Manual*. N. Bethesda, Maryland: Second Language Testing, Inc.

Ramus, F. (2002a). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition, 2*, 85–115.

Ramus, F. (2002b). Acoustic correlates of linguistic rhythm: Perspectives. *Proceedings of Speech Prosody 2002*, Aix-en-Provence (pp. 115–120). Aix-en-Provence: Laboratoire Parole et Langage.

Russo, F., Windell, D., Weinert, S., & Cuddy, L. (2003). Learning the “special note”: Evidence for a critical period for absolute pitch acquisition. *Music Perception, 21*(1), 119–127.

Sacks, O. (2007). *Musicophilia: Tales of music and the brain*. New York, NY: Knopf.

Satoi, H. (2003). Mora-timing rhythm in Japanese English. In S. Watanabe (Ed.), *Onsei gengo no kenkyu* (pp. 19–28). Osaka: Graduate School of Language and Culture, Osaka University.

Schellenburg, E. G. (2003). Does exposure to music have beneficial side effects? In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 430–448). Oxford, UK: Oxford University Press.

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

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of Acoustical Society of America, 71,* 996–1007.

Seashore, C. B., Lewis, C., & Saaeteweit, J. G. (1960). *Seashore measures of musical talent:*

*Manual.* San Antonio, TX: Psychological Corporation.

Selkirk, E. (1978). *On prosodic structure and its relation to syntactic structure* (Doctoral dissertation). University of Massachusetts, Amherst.

Shahin, A., Roberts, L., & Trainor, L. (2004). Enhancement of auditory cortical development by musical experience in children. *Neuro Report, 15*, 1919–1921.

Shimizu, T. (2002). The effect of listening to a familiar music piece to a pregnant mother and her unborn baby. *The 43rd Asahikawa Conference of the Japan Society of Maternal Health.*

Slavoff, G., & Johnson, J. (1995). The effects of age on the rate of learning a second language. *Studies in Second Language Acquisition, 17*, 1–16.

Slevc, L. R., & Miyake, A. (2006). Individual difference in second-language proficiency: Does musical ability matter? *Psychological Science, 17*(8), 675–681.

Steinhauer, K., Alter, K., & Friederici, A. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*, 191–196.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin, 113*, 345–361.

Takeuchi, O., & Mizumoto, A. (2012). Gaikokugo kyoiku kenkyu handbook. A handbook of research in foreign language learning and teaching. Tokyo, Japan: Shohakusha.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech effects of music lessons. *New York Academy of Science, 999,* 530–532. doi: 10.1196/annals.1284.067

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4,* 46–64.

Trainor, L. J. (2005). Are there critical periods for musical development? *Developmental Psychobiology, 46,* 262–278. doi: 10.1002/dev.20059

Turnbull, M., Lapkin, S., Hart, D., & Swain, M. (1998). Time on task and immersion graduates' French proficiency. In S. Lapkin (Ed.), *French second language education in Canada: Empirical studies* (pp. 31–55). Toronto, Canada: University of Toronto Press.

Upshur, J., Spaan, M., & Thrasher, R. (1969). *Michigan test of aural comprehension*. Ann Arbor: University of Michigan.

Vergnaud, R., & Halle, M. (1979). *Metrical structures in phonology – A fragment of a draft*. Mimeographed. MIT, Cambridge, Massachusetts.

Warner, N., & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America, 109*(3), 1144–1156.

Weinert, S. (1992). Deficits in acquiring language structure: The importance of using prosodic cues. *Applied Cognitive Psychology, 6,* 545–571.

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford, UK: Oxford University Press.

Wing, H. D. (1968). *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity*. London, UK: Cambridge University Press.

Yusawa, M., Sekiguchi, M., & Rie, S. (2007). Phonological awareness of English among young Japanese children. *The Bulletin of Graduate School of Education of Hiroshima University, 3*(56), 153–160.

# Appendices

## Appendix A1 Chunking Input Test Script

| | **Items** | **Cue** |
|---|---|---|
| **1** | Black and pink & red | 1 |
| **2** | Chicken-fingers and fruit | 2 |
| **3** | Black & pink and red | 2 |
| **4** | Cream-buns & chocolate | 2 |
| **5** | Pink and black & green | 1 |
| **6** | Black and green & pink | 1 |
| **7** | Chocolate cookies and jam | 1 |
| **8** | Pink & black and green | 2 |
| **9** | Black & green and pink | 2 |
| **10** | Chocolate ice cream & honey | 2 |
| **11** | Fruit salad & milk | 1 |
| **12** | Black and pink & green | 1 |
| **13** | Chocolate ice cream and buns | 1 |
| **14** | Fruit-salad and milk | 2 |
| **15** | Chocolate cookies and jam | 2 |
| **16** | Red and pink & black | 1 |
| **17** | Red and pink & black | 1 |
| **18** | Cream buns & chocolate | 2 |

Appendix A2 Focus Input Test Script

| | **Items** | **Cue** |
|---|---|---|
| **1** | I wanted BLACK & red socks. | 1 |
| **2** | I wanted blue & GREEN socks. | 2 |
| **3** | I wanted red & BLACK socks. | 2 |
| **4** | I wanted BLUE & red socks. | 1 |
| **5** | I wanted GREEN & blue socks. | 1 |
| **6** | I wanted green & BLACK socks. | 2 |
| **7** | I wanted blue & RED socks. | 2 |
| **8** | I wanted GREEN & black socks. | 1 |
| **9** | I wanted RED & blue socks. | 1 |
| **10** | I wanted BLACK & green socks. | 1 |
| **11** | I wanted green & BLUE socks. | 2 |
| **12** | I wanted RED & black socks. | 1 |
| **13** | I wanted black & GREEN socks. | 2 |
| **14** | I wanted BLUE & green socks. | 1 |
| **15** | I wanted BLUE & green socks. | 2 |
| **16** | I wanted red & BLUE socks. | 2 |

Appendix A3 Prosody Discrimination Test Script

1. (prosody of Chinese and English in female voice)

   (English) Our teacher gave us chocolate on Valentine's Day.

   (Chinese) 我们的老师在情人节送给了我们巧克力。

2. (prosody of English and Russian in male voice)

   (English) Why do you think that she likes your best friend?

   (Russian) Все члены семьи живут вместе в маленьком доме.

3. (prosody of English and French in male voice)

   (English) Is Jack an elementary school student?

   (French) Est-ce que Jack est un élève d'école primaire?

4. (prosody of Arabic and English in male voice)

   (English) All family members live together in a small house.

   الاسرة تعيش فى منزل صغير.

5. (prosody of English and Polish in male voice)

   (English) Our teacher gave us chocolate on Valentine's Day.

   (Polish) Nasza nauczyciela dała nam czekoladę na Walentynki.

6. (prosody of English and Korean in female voice)

   (English) All family members live together in a small house.

   (Korean) 우리가족 모두는 조그만 집에서 함께 삽니다.

7. (prosody of English and German in male voice)

   (English) My sister Jane goes shopping in town every Friday.

   (German) Meine Schwester Jane geht jeden Freitag in die Stadt zum Einkaufen.

8. (prosody of English and French in male voice)

    (English) Everyone in class knows the reason for their quarrel.

    (French) Tout le monde dans la classe connaît la raison de leur querelle.

9. (prosody of English and Russian in male voice)

    (English) Bring me some water right now! I am so thirsty.

    (Russian) Скорее принеси мне воды! Очень хочется пить.

10. (prosody of English and Japanese in female voice)

    (English) My sister Jane goes shopping in town every Friday.

    (Japanese) 私の姉のジェーンは、毎週金曜に街へ買い物に行きます。

11. (prosody of English and German in female voice)

    (English) Our teacher gave us chocolate on Valentine’s Day.

    (German) Unser Lehrer gab uns Schokolade am Valentinstag.

12. (prosody of English and Mongolian in female voice)

    (English) My favorite food is soup curry and I cook it myself once a month.

    (Mongolian) Би сүүп карри-нд их дуртай, сардаа нэг удаа хийж иддэг шүү.

# Appendix A4

## An Answer sheet for the Prosody Discrimination Test

「英語はどちら？」　解答用紙

学年　　　組　名前

☆これから読まれる２つの文のうち、英語文はどちらでしょうか？最初（１番目）に読まれた方ならば、１に○を、２番目に読まれた方ならば、２に○をつけなさい。

解答の仕方　　①　　2

| 英語ともう一つは・・・ | 英語だと思うのは・・・ | |
|---|---|---|
| 問題１　中国語 Mandarin Chinese | 1 | 2 |
| 問題２　ロシア語 Russian | 1 | 2 |
| 問題３　フランス語 French | 1 | 2 |
| 問題４　アラビア語 Arabic | 1 | 2 |
| 問題５　ポーランド語 Polish | 1 | 2 |
| 問題６　韓国語 Korean | 1 | 2 |
| 問題７　ドイツ語 German | 1 | 2 |
| 問題８　フランス語 French | 1 | 2 |
| 問題９　ロシア語 Russian | 1 | 2 |
| 問題 10　日本語 Japanese | 1 | 2 |
| 問題 11　ドイツ語 German | 1 | 2 |
| 問題 12　モンゴル語 Mongolian | 1 | 2 |

Question for 1-12: which is English?

Answer1: 2, A2: 1, A3: 1, A4: 2, A5:1, A6: 2, A7:1, A8: 1, A9:1, A10:1, A11:2, A12: 2.

# Appendix B

## Eiken Grade 3 Listening Test Question Samples

1st part (conversation sentences and answer choices for continuous sentences)

Picture (attached in the question sheet): in the picture, Annie and a boy are talking in the kitchen.

Sentences : I am hungry, Annie.

: Me, too. Let's make something.

: How about pancakes?

Answers: 1 On the weekend. 2 For my friends. 3 That's a good idea.

The correct answer is 3.

2nd part (conversation sentences and answer choices for continuous sentences)

Sentences : Do you have any pets?

: Yes, Linda. I have two dogs and a cat.

: You are lucky. I only have a bird.

: You should come and play with my pets.

Question : How many pets does Linda have?

Answers : 1 One. 2 Two. 3 Three. 4 Four.

The correct answer is 1.

3rd part (passage and answer choices for continuous sentences)

Passage : The weather was terrible during my school trip. On Monday and Tuesday, it was cloudy and windy. On Wednesday, it snowed. Finally, it was sunny when we came home on Thursday.

Question : How was the weather on Thursday?

Answers: 1 Windy. 2 Cloudy. 3 Sunny. 4 Snowy.

The correct answer is 3.

# Appendix C

## INTERMEDIATE MEASURES OF MUSIC AUDIATION

## TONAL TEST

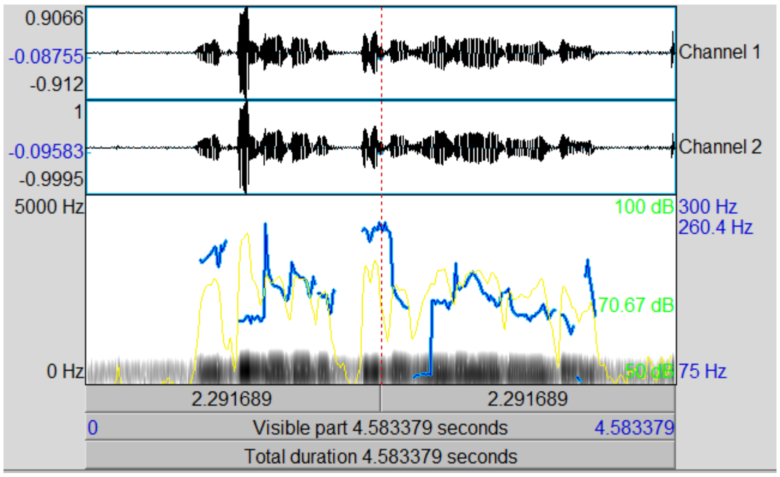

*Note: The participants choose ☺ ☺ for the same tonal pair, and choose ☺ ☹ for different ones, in their answer sheets. The model answers for the practice examples are: 1 ☺ ☺, 2 ☺ ☹, 3☺ ☹, and 4 ☺ ☺.*

# Appendix D


*Note: The participants choose ☺ ☺ for the same rhythm pair, and choose ☺ ☹ for different ones, in their answer sheets. The model answers for the practice examples are: 1 ☺ ☹ and 2 ☺ ☺.*

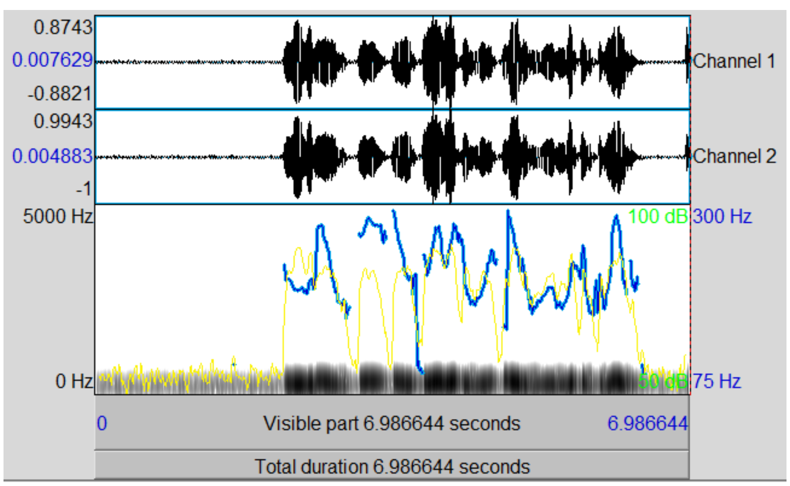
# Appendix E

## Prosody Discrimination “Praat” Graphs (Boersma, 2001)



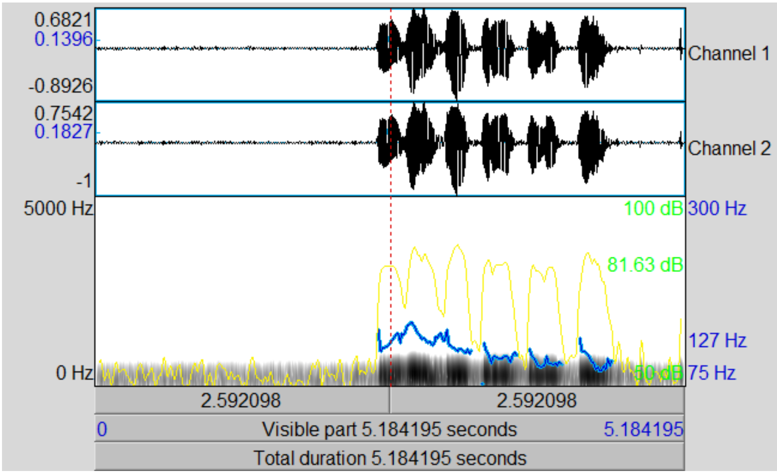

Prosody test Q1 American English female

(English) Our teacher gave us chocolate on Valentine’s Day.

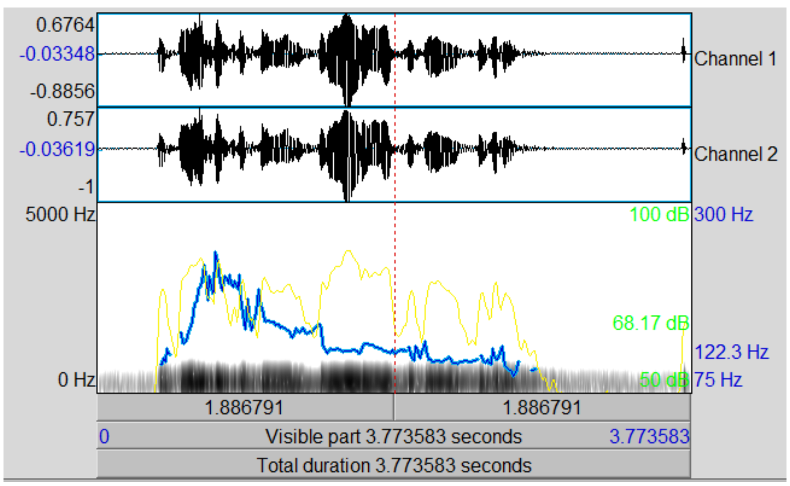



Prosody test Q1 Chinese female

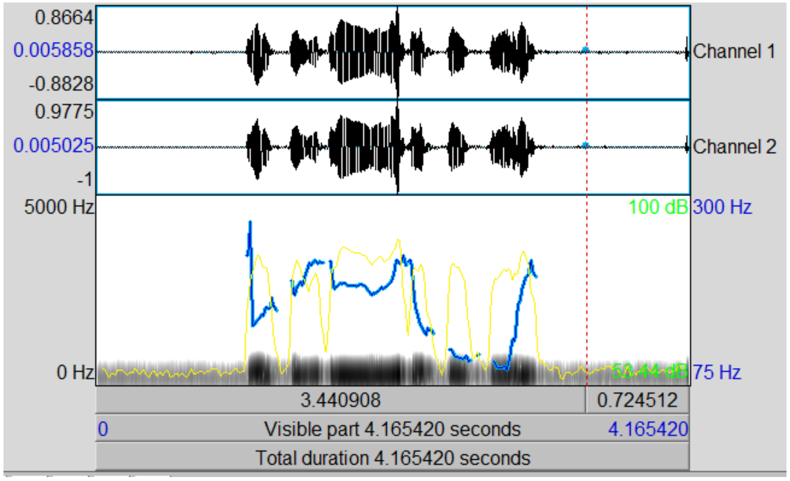
(Chinese) 我们的老师在情人节送给了我们巧克力。



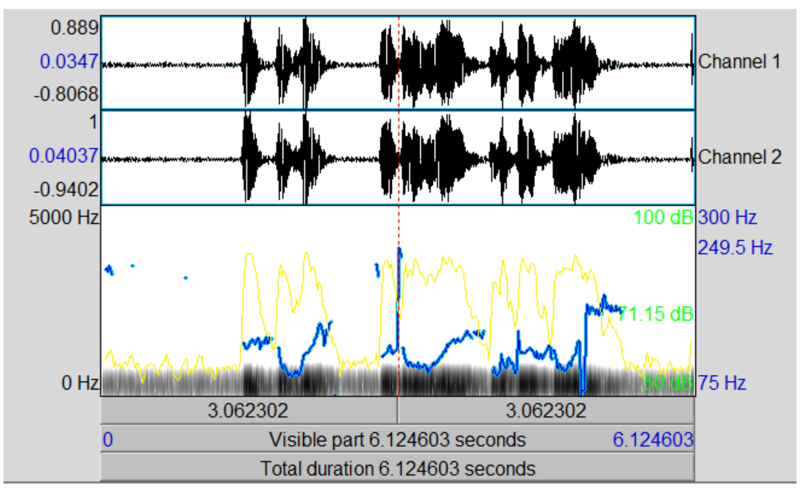

Prosody test Q2 American English male
(English) Why do you think that she likes your best friend?



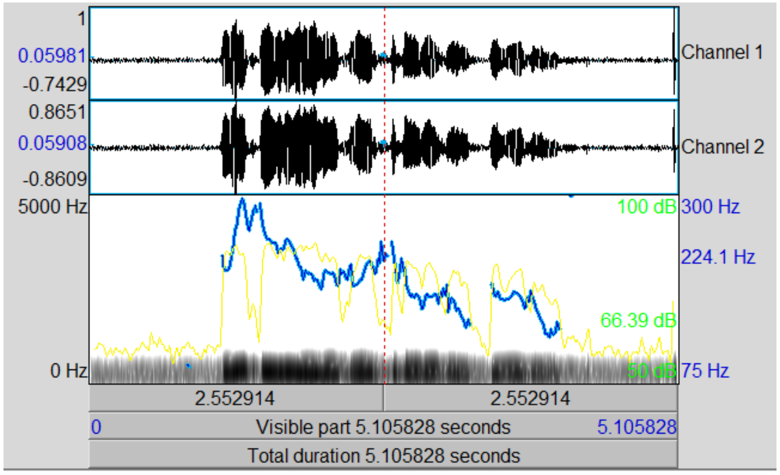

Prosody test Q2 Russian male
(Russian) Все члены семьи живут вместе в маленьком доме.



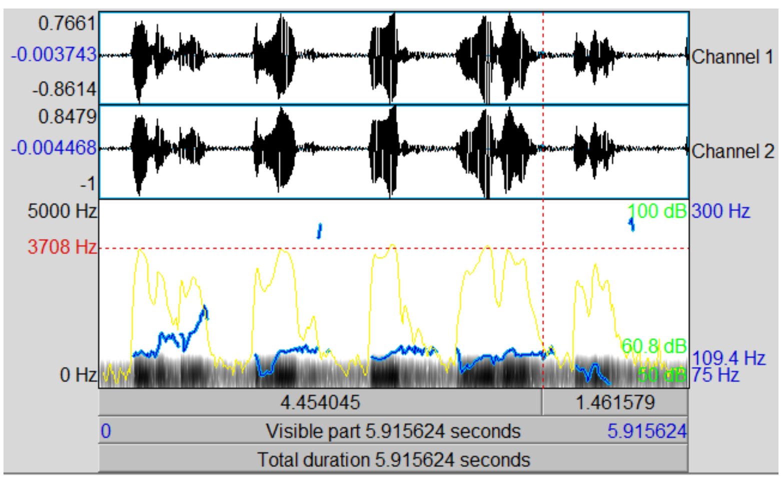

Prosody test Q3 Australian male
(English) Is Jack an elementary school student?



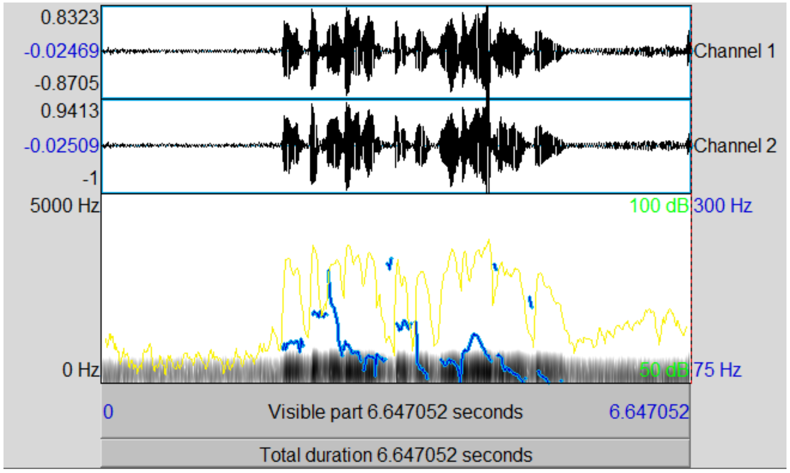

Prosody test Q3 French male
(French) Est-ce que Jack est un élève d’école primaire?




Prosody test Q4 American male
(English) All family members live together in a small house.

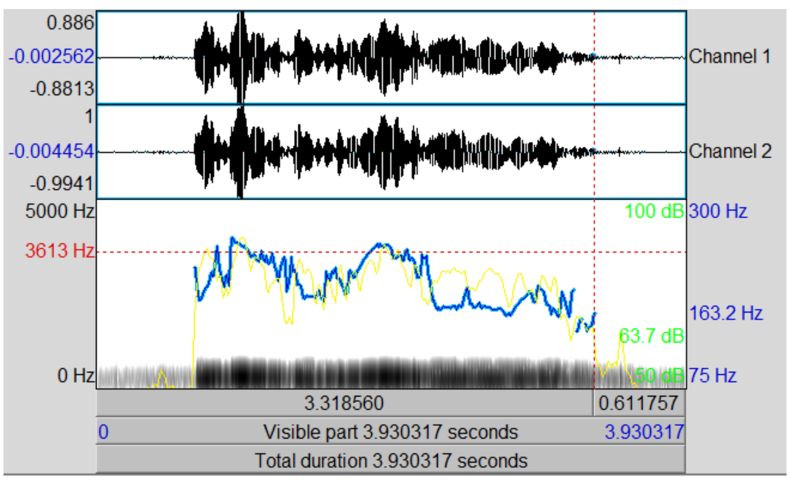



Prosody test Q4 Arabic male
الاسرة تعيش فى منزل صغير.



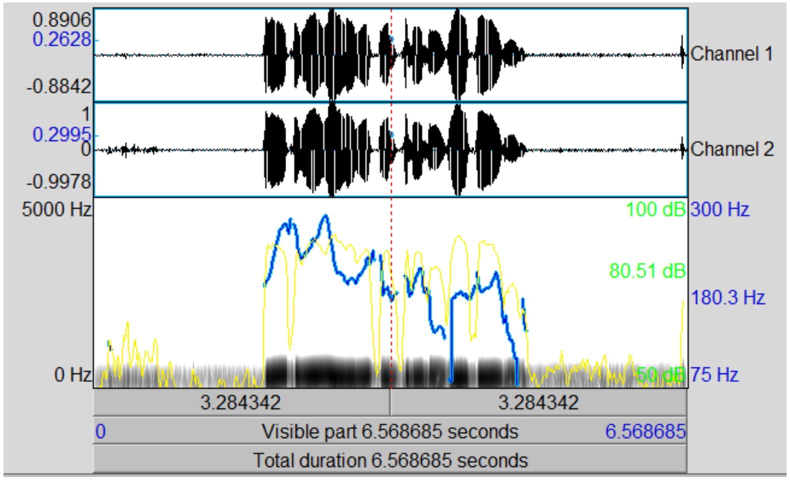

Prosody test Q5 American male

(English) Our teacher gave us chocolate on Valentine's Day.

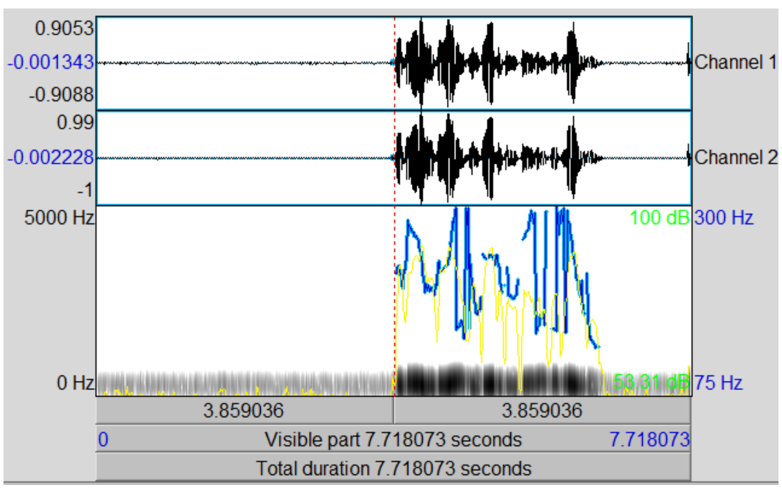



Prosody test Q5 Polish male

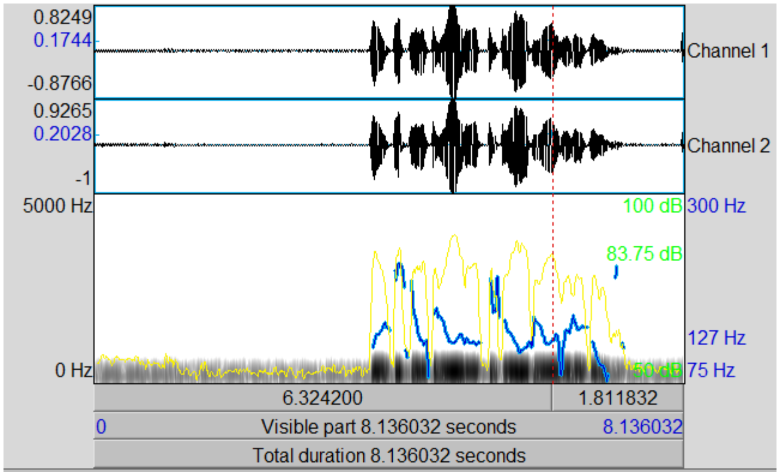
(Polish) Nasza nauczyciela dała nam czekoladę na Walentynki.




Prosody test Q6 English spoken by Japanese female
(English) All family members live together in a small house.

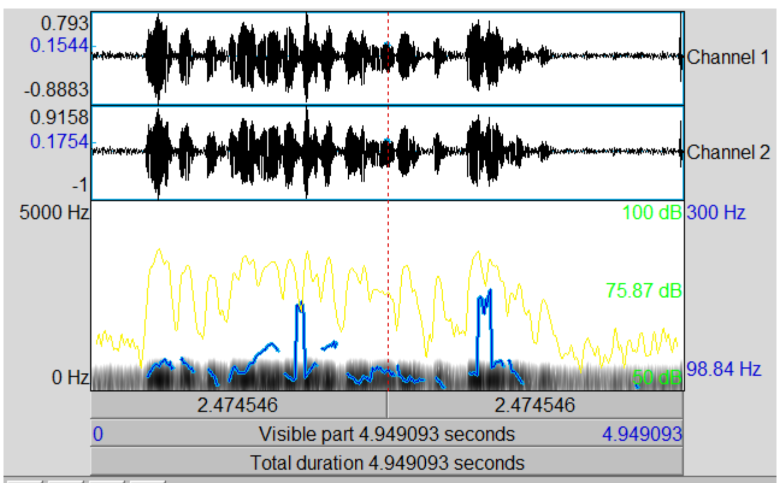



Prosody test Q6 Korean female
(Korean) 우리가족 모두는 조그만 집에서 함께 삽니다.



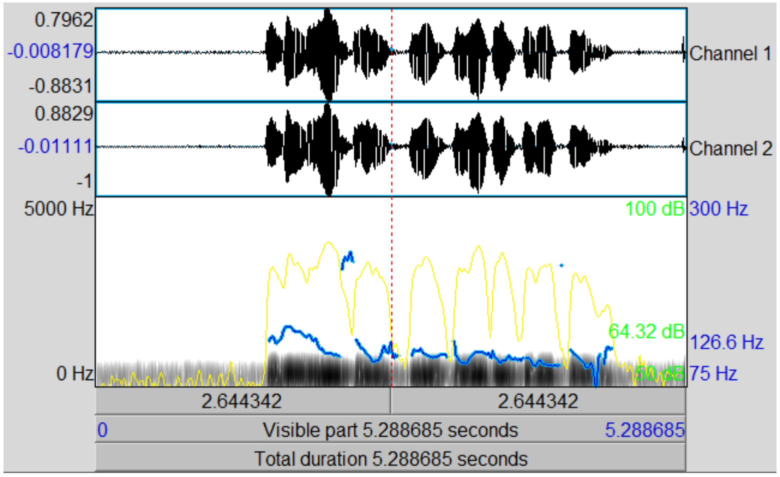

Prosody test Q7 Australian male

(English) My sister Jane goes shopping in town every Friday.

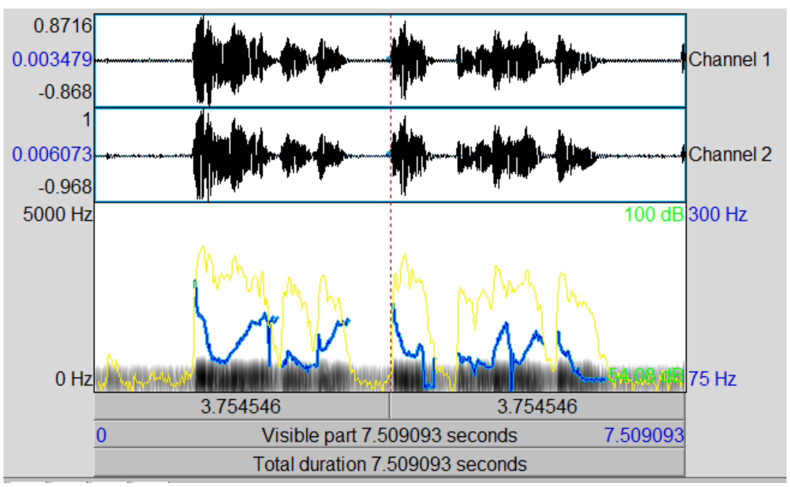



Prosody test Q7 German male

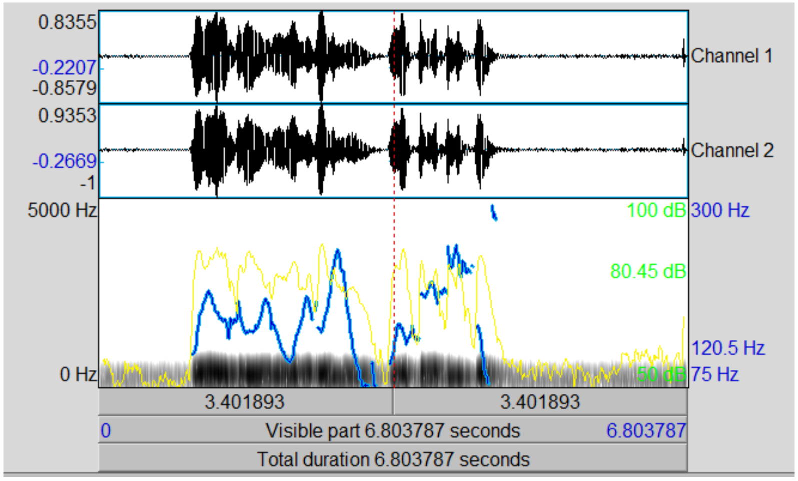
(German) Meine Schwester Jane geht jeden Freitag in die Stadt zum Einkaufen.



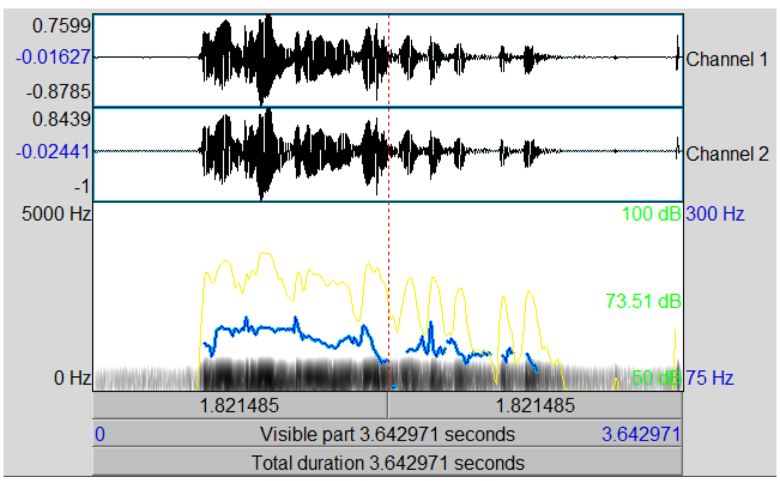

Prosody test Q8 American male
(English) Everyone in class knows the reason for their quarrel.



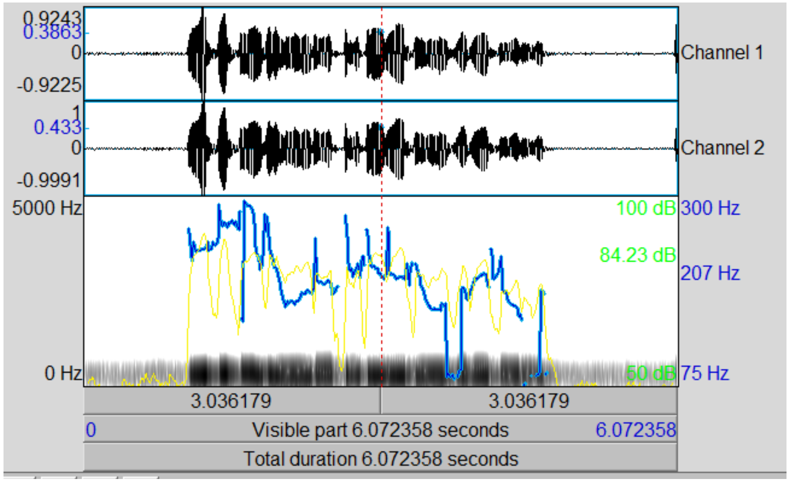

Prosody test Q8 French male
(French) Tout le monde dans la classe connait la raison de leur querelle.




Prosody test Q9 American male
(English) Bring me some water right now! I am so thirsty.

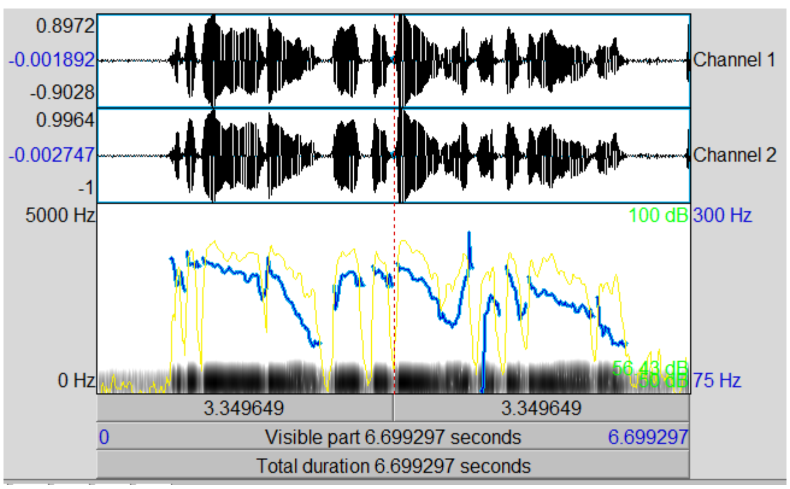



Prosody test Q9 Russian male
(Russian) Скорее принеси мне воды! Очень хочется пить.



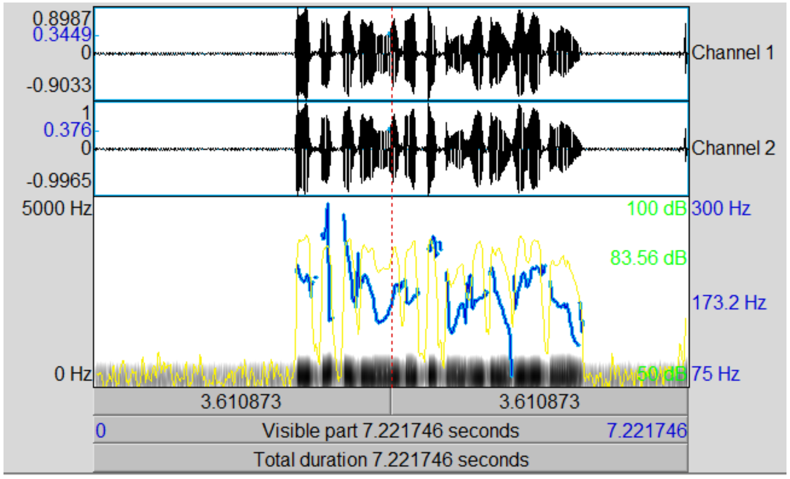

Prosody test Q10 American female

(English) My sister Jane goes shopping in town every Friday.

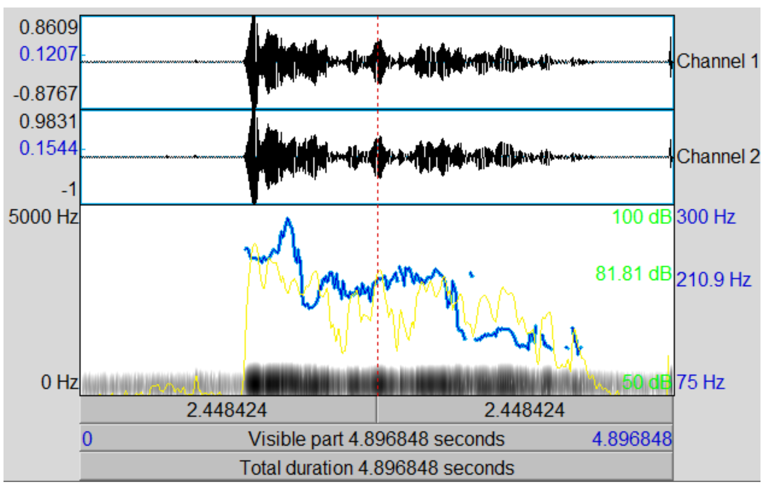



Prosody test Q 10 Japanese female

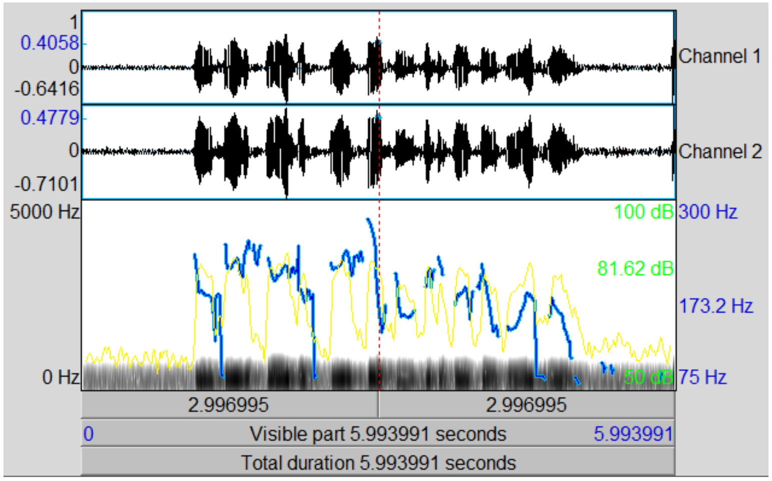
(Japanese) 私の姉のジェーンは、毎週金曜に街へ買い物に行きます。




Prosody test Q11 English spoken by Japanese female
(English) Our teacher gave us chocolate on Valentine's Day.

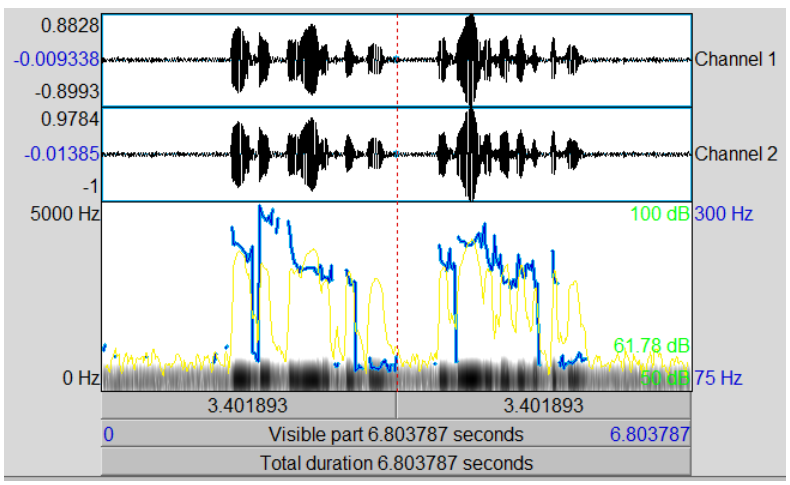



Prosody test Q11 German female
(German) Unser Lehrer gab uns chokolade am Valentinstag.




Prosody test Q12 American female

(English) My favorite food is soup curry and I cook it myself once a month.




Prosody test Q12 Mongolian female

(Mongolian) Би сүүп карри-нд их дуртай, сардаа нэг удаа хийж иддэг

# Appendix F

## Study 1 Questionnaire on Music and EFL for High School Students

高校生英語・音楽アンケート

☺ 以下のアンケートにお答えください。

1 今、何歳何か月ですか？

2 音楽を学校以外で習っていますか？何をならっていますか（歌？どんな楽器？）

3 音楽を学校以外で習い始めたのは何歳ですか？

4 英語を学校以外で習い始めたのは何歳何か月ですか？

5 学校で英語を習い始めたのは、何歳のときですか？

6 楽器・歌は毎日何時間何分練習しますか？

7 英語は授業以外で、週に合計、何時間何分勉強しますか？

8 英語を使うのが好きですか？

9 音楽を演奏するのが好きですか？

10 楽譜が読めますか？

11 絶対音感がありますか？（音を聴いて、その音の高さ「ドレミファソラシド」が分かる）

12 簡単な英単語・英文を読むことができますか？

13 簡単な英単語・英文を書くことができますか？

14 簡単な英単語・英文を話すことができますか？

15 簡単な英単語・英文を聴いて理解することができますか？

16 高校で週に何時間英語の授業がありますか？

English translation of the questionnaire in Appendix F

1-How old are you?

2-Do you take music lessons, including musical instruments and choir, outside of school?

3-When did you start learning music outside of school?

4-When did you start learning English outside of school?

5-When did you start learning English at school?

6-How long do you practice musical instruments or singing per day or per week?

7-How long do you study English outside of school per week?

8-Do you like using English?

9-Do you like performing music?

10-Can you read music scores or notes?

11-Do you have absolute pitch?

12-Can you read simple English words or sentences?

13-Can you write simple English words or sentences?

14-Can you speak simple English words or sentences?

15-Can you comprehend simple English words or sentences?

16-How many hours do you study English at school per week?

## Appendix G

Table G1

*Independent t-Test on Tonal Test Scores of Male and Female High School Students*

| | HS Male | HS female |
|---|---|---|
| *N* | 16 | 59 |
| *M* | 27.40 | 33.80 |
| *SD* | 12.70 | 6.40 |
| *d* | -.79 | |
| *t*(73) | 1.95 (*ns*) | |

*Note:* Equal variances were not assumed; *d* indicates effect size, i.e., Cohen's *d*.

## Appendix H

Table H1

*Independent t-Test on Rhythm Scores of Male and Female High School Students*

| | HS Male | HS female |
|---|---|---|
| *N* | 16 | 59 |
| *M* | 29.00 | 32.50 |
| *SD* | 11.00 | 6.20 |
| *d* | -.47 | |
| *t*(73) | 1.22 (*ns*) | |

*Note:* Equal variances were not assumed; *d* indicates effect size, i.e., Cohen's *d*.

## Appendix I

Table I1

*Independent t-Test on Eiken Grade 3 Listening Test Scores of Students in Their First and Second Years of High School*

| | Year 1 students | Year 2 students |
|---|---|---|
| *N* | 19 | 56 |
| *M* | 12.90 | 11.80 |
| *SD* | 3.80 | 4.50 |
| *d* | .25 | |
| *t*(73) | .96 (*ns*) | |

*Note: d* indicates effect size, i.e., Cohen's *d*.

# Appendix J

Table J1

*Spearman's Correlation Matrix between the Musical Sound Sensitivity Tests, English Prosody Tests, and English Listening Test*

| Variable | Prosody input | Focus input | Chunking input | Prosody discrimination | Eiken Grade 3 listening test | Tonal test | Rhythm test |
|---|---|---|---|---|---|---|---|
| Prosody input | 1 | | | | | | |
| Focus input | .33** | 1 | | | | | |
| Chunking input | .18 | .32** | 1 | | | | |
| Prosody discrimination | .049 | .022 | -.13 | 1 | | | |
| Eiken Grade 3 listening test | .29* | .37** | .29* | .092 | 1 | | |
| Tonal test | .12 | .15 | -.045 | .14 | .087 | 1 | |
| Rhythm test | .23* | .086 | -.056 | .12 | .23* | .53*** | 1 |

*Note:* *$p < .05$, **$p < .01$, ***$p < .001$.

Spearman's correlation analysis revealed weaker correlation between tone and rhythm processing skills, English prosody processing skills, and English listening comprehension. Tone processing skill did not correlate with any English prosody processing skill. Rhythm processing skill significantly correlated with prosody input ($r = .23$, $p < .05$).

Spearman's correlation analysis revealed a weak correlation between the Eiken Grade 3 listening test score and prosody input ($r = .29$, $p < .05$), Eiken Grade 3 listening test score and focus input ($r = .37$, $p < .01$), Eiken Grade 3 listening test score and chunking input ($r = .29$, $p = .05$), and Eiken Grade 3 listening test score and rhythm processing skill ($r = .23$, $p < .05$).

# Appendix K

## Scoring Sheet for the Prosody Production Test and the Rhythm Production Test

| | Prosodic production<br>G (good), F (fair), or P (poor) | Rhythmic Production<br>G (good), F (fair), or P (poor) | | Total score of Rhythm production test |
|---|---|---|---|---|
| 1 | Black, and green & pink socks | 1st time | 2nd time | |
| 2 | Black and green socks | | | |
| 3 | Pink & red, and black socks | | | |
| 4 | Black and green socks | | | |
| 5 | Chocolate cookies, and jam | | | |
| 6 | Black & pink, and red socks | | | |
| 7 | Cream, buns, and chocolate | | | |
| 8 | Red and black socks | | | |
| 9 | Blue and red socks | | | |
| 10 | Red and black shoes | | | |
| 11 | Chocolate-ice-cream, and honey | | | |
| 12 | Red, and black & pink socks | | | |
| 13 | Blue and red socks | / | | / |
| 14 | Black and green shoes | / | | / |
| 15 | Cream-buns and chocolate | / | | / |
| 16 | Red and black socks | / | | / |
| In total | | | | |

# Appendix L

## Mother Goose Notes for the Rhythm Production Test

1. Mary Had a Little Lamb
2. Lavender’s Blue
3. Hush, Little Baby
4. What Are Little Boys Made Of?
5. Polly Put the Kettle On
6. Hush-a-Bye, Baby
7. Sally, Go Round
8. Old MacDonald Had a Farm
9. Row, Row, Row Your Boat
10. Head, Shoulders, Knees, and Toes
11. This Old Man
12. London Bridge Is Falling Down

# Mother goose

10

11

12

# Appendix M

## Sample Questions from JET 7–8 Levels

Parts 1-2 ask for vocabulary definitions.
<Part 1 instruction> Listen and choose the correct picture (from the three pictures in the question sheet).

(Script for the sample question) A comb

<Part 2 instruction> Listen and choose the correct word (which describes the picture).

(Script) (1) A painter (2) A farmer (3) A doctor

Parts 3–4 ask for sentence meanings.
<Part 3 instruction> Listen and choose the correct picture (describing three choices of clocks showing different times).
(Script) It's eight o'clock.

<Part 4 instruction>Listen and choose the correct sentence (describing the picture).
(Script) The girl is eating cookies. (2) The girl is buying cookies. (3) The girl is baking cookies.

Part 5 asks for conversation meaning.
<Part 5 instruction> Listen and choose the correct picture (from three pictures describing different situations of the woman).

(Script) Woman: Jessie! Where are you? Girl: I am in the bathroom.

Part 6 asks for communication skill.
<Part 6 instruction> Listen and choose the correct response. Mark the correct answer on your answer sheet.

(Script) Happy birthday. This is for you. (1) Thank you. (2) I'm sorry. (3) It's on the table.

Part 7 asks for written word meanings.

<Part 7 instruction> Observe and choose the correct word (describing the picture).

(Written choices) (1) A bed (2) A table (3) A lamp

## Appendix N

### Study 2 Music Questionnaire for Elementary School Students

小学生音楽アンケート

当てはまるほうに○をつけたり、（　　　　　　　　）に答えてください。

1・いま、音楽（うたや楽器）を学校のじゅぎょうのほかに、ならっていますか？

（はい　・　いいえ　・　いまはならっていないが、前にならっていた）

2・いつからならっていますか？またはならっていましたか？

（　　　　　　　　　　さい　から）

3・なにをならっていますか？うたですか？どんながっきですか？

（　　　　　　　　　　　　　　）

4・がくふ（おんぷ）が自分でよめますか？

（　よめる　・　すこしはよめる　・　ぜんぜんよめない）

5・うたか楽器をどれくらいれんしゅうしますか

（　毎日 ________分練習する　・　______日に一回、______分練習する

あまり練習しない　・　ぜんぜんれんしゅうしない　）

6・うたうことや楽器を演奏することがすきですか？（　好き　・　ふつう　・　好きでない）

7・うたや楽器の演奏をきくことがすきですか？　（　好き　・　ふつう　・　好きでない）

English translation of the questionnaire in Appendix N

1-Do you take music lessons, including instrumental music and choir, outside of school?
(Yes. / No. / I did previously.)

2-When did you start taking music lessons outside of school? (Since I was years old.)

-When did you start taking music lessons outside of school? (From years old to years old.)

3-What music do you learn outside of school? (I learn .)

4-Can you read musical scores or musical notes? (Very well. / A little. / Not at all.)

5-How long do you practice music per day or per week?
( hours minutes every day. / hours minutes per . / I do not practice.)

6-Do you like performing music? (Yes. / So so. / No.)

7-Do you like listening to music? (Yes. / So so. / No.)

# Appendix O

## Study 2 EFL Questionnaire for Elementary School Students

小学生英語アンケート

当てはまるほうに○をつけたり、（　　　　　　）に答えてください。

1・学校の授業のほかに英語をならっていますか？

（　はい　・　いいえ　）

2・いつから習っていますか？　（　　　　さい　から　）

3・いまは英語をならっていないが、前に習っていましたか？

（　はい　・　いいえ　）

4・英語をはなすことができますか？

（　上手にできる　・　ふつう　・　できない　）

5・英語をきいて、何をいっているか分かりますか

（　よく分かる　・　半分わかる　・　少しわかる　・　ぜんぜん分からない　）

6・英語で書くことができますか？

（　文を書くことができる　・　みじかいことばを書ける　・　アルファベットをかける　）

7・英語で読むことができますか？

（　お話を読むことができる　・みじかい文を読める　・文は読めないが、ことばは読める　）

8・英語をまなぶことは楽しいですか？（　たのしい　・　ふつう　・　たのしくない）

9・なにか、ききたいことや感じたことがあったら書いてください

（　　　　　　　　　　　　　　　　）

English translation of the questionnaire in Appendix O

1-Do you learn English outside of school? (Yes. / No.)

2-When did you start learning English outside of school? (Since I was years old.)

3-Did you previously learn English outside of school? (Yes, when I was years old. / No.)

4-Can you speak English? (Very well. / So so. / No.)

5-Can you understand spoken English? (Very well. / So so. / A little. / Not at all.)

6-Can you write in English? (I can write sentences. / I can write words. / I can write letters.)

**7-**Can you read English? (I can read stories. / I can read short sentences. / I can read words.)

8-Do you enjoy learning English? (Yes. / So so. / No.)

9-If you have any questions or comments, please ask.

# Appendix P

## Statistical Results from the Item-Trimmed Prosody Discrimination Test

Table P1

*Descriptive Statistics for the Tests Used in This Study*

| Variable | *K* | *M* | *SD* | Range | *α* | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| **Prosodic skill measurements:** | | | | | | | |
| Prosody discrimination (item-trimmed) | 8 | 4.23 | 2.08 | 0-8 | .613 | -.27 | -.84 |

*Note: K* indicates the number of items in each test.

Table P2

*Pearson's Correlation Matrix between the Musical Sound Perception/Production Tests and the English Prosody Test, i.e., Revised Prosody Discrimination Test*

| Variable | Prosody input | Focus input | Chunking input | Prosody discrimination | Prosody production | Tone | Rhythm | Productive rhythm | JET |
|---|---|---|---|---|---|---|---|---|---|
| Prosody discrimination (item-trimmed) | -.067 | .18 | .013 | .90*** | .26 | .28* | .019 | .20 | .004 |

*Note:* *$p < .05$, ***$p < .001$.

Table P3

*Collective Multiple Regression Analysis with JET Score as a Dependent Variable and English Prosody Perception/Production as Independent Variables*

| Predictor | B | *SEB* | *Β* |
|---|---|---|---|
| Prosody input | .26 | .20 | .13 |
| Focus input | .056 | .39 | .014 |
| Chunking input | .35 | .18 | .20 |
| Prosody discrimination (item-trimmed) | -3.96 | .24 | -.17 |
| Prosody production | .72 | .12 | .64*** |
| Intercept | 11.01 | 5.70 | |

*Note:* $R^2$ = .60, ****p* < .001.

English prosody perception/production accounted for 60% of JET score.

## Appendix Q

Table Q1

*Inter-rater Reliability of Prosody Production Test Scores between Two Raters*

| | $\alpha$ |
|---|---|
| Prosody production test | .98 |

*Note: N = 53*, the raters' mean scores: JET (*M* = 6.64), AET (*M* = 6.92).

# Appendix R

Table R1

*Inter-rater Reliability of Rhythm Production Test Scores*

| | *α* |
|---|---|
| Rhythm production test | .96 |

*Note: N = 53*, the raters' mean scores: violin student (*M* = 7.21), professional pianist (*M* = 7.47).

## Appendix S

Table S1

*Independent t-Analysis of Tonal Test Scores of Male and Female Elementary School Students*

| | Male | Female |
|---|---|---|
| *N* | 27 | 26 |
| *M* | 35.50 | 34.50 |
| *SD* | 2.60 | 4.70 |
| *d* | | .26 |
| *t*(51) | | .95 (*ns*) |

*Note:* Equal variances were not assumed; *d* indicates effect size, i.e., Cohen's *d*.

## Appendix T

Table T1

*Independent t-Analysis of Rhythm Test Scores of Male and Female Elementary School Students*

| | Male | Female |
|---|---|---|
| *N* | 27 | 26 |
| *M* | 31.50 | 31.90 |
| *SD* | 1.10 | 6.20 |
| *d* | -.09 | |
| *t*(51) | .32 (*ns*) | |

*Note:* Equal variances were not assumed; *d* indicates effect size, i.e., Cohen's *d*.

# Appendix U

## Study 4 Questionnaire for Mothers of Musically Trained/Untrained Children

音楽学習者　保護者アンケート

【アンケート】保護者の方、回答にご協力お願いします。

1 今、何歳何か月ですか？

2 習っている楽器は何ですか？

3 習い始めたのは何歳ですか？

4 楽器は毎日何時間何分練習しますか？

5 音楽を演奏するのが好きですか？

6 楽譜が読めますか？

7 絶対音感がありますか？（音を聴いて、ドレミファソラシドが分かる）

English translation of the Questionnaire in Appendix U

1-How old is your son or daughter?

2-What musical instrument does your son or daughter practice daily?

3-When did your son or daughter start practicing the musical instrument?

4-For how long does your son or daughter practice the musical instrument per day?

5-Does your son or daughter like playing music?

6-Can your son or daughter read musical scores?

7-Does your son or daughter have absolute pitch?

# Appendix V

Regression Tree Analysis Results of the Prosody Discrimination Test of High School Participants

The analysis revealed that the language prosody type did not contribute to the Prosody discrimination test scores of the high school participants. The nine languages were grouped into nodes differently according to the hypothesized language prosody types of stress-timed, syllable-timed, and mora-timed. The regression tree is sub-structured with the first node comprising Arabic and Japanese as the most contributing variables; the second node of French, Polish, German, and Mongolian as the second-most contributing variables; and the third node of Chinese, Russian, and Korean as the third-most contributing variables.

| category | % | $N$ |
|---|---|---|
| Wrong answer | 31.1 | 210 |
| Right answer | 68.9 | 465 |
| Sum | 100 | 675 |

Arabic, Japanese

Node1

| category | % | $N$ |
|---|---|---|
| Wrong answer | 17.3 | 26 |
| Right answer | 82.7 | 124 |
| sum | 22.2 | 150 |

French, Polish, German, Mongolian

Node2

| category | % | $N$ |
|---|---|---|
| Wrong answer | 29.3 | 88 |
| Right answer | 70.7 | 212 |
| sum | 44.4 | 300 |

Chinese, Russian, Korean

Node 3

| category | % | $N$ |
|---|---|---|
| Wrong answer | 42.7 | 96 |
| Right answer | 57.3 | 129 |
| sum | 33.3 | 225 |

*Notes:* $X^2 = 27.75$, $p = .003$, $df = 2$. *N* indicates the number of right or wrong answers.

## Endnotes

Chapter 1 Introduction

[1] *Eiken*: Eiken Test in Practical English Proficiency, Listening Test (Test of English as a Foreign Language).

The Eiken Foundation of Japan, as Japan's largest testing body, produces and administers English proficiency tests with the backing of the Japanese Ministry of Education, Culture, Sports, Science and Technology and in cooperation with Japanese prefectural and local boards of education, public and private schools, and other leading testing bodies. The Eiken tests the general skills of English, with a compulsory speaking test.

[2] *audiation*: "[I]t is the process that takes place when we hear and comprehend music for which the sound is no longer physically present. In contrast, aural perception takes place when we simply hear sound that is physically present. Sound is not comprehended as music until it is audiated after it is heard. We may audiate while listening to music, recalling, performing, interpreting, creating, improvising, reading, and writing music notation…. As we listen to the music, we aurally perceive sound the moment that it is heard. It is not until a moment or so after the sound is heard that we audiate and give meaning to that sound while aurally perceiving and giving meaning to additional sounds that will follow in the music" (Gordon, 1994, p. 40).

Chapter 2 Literature Review

[3] *absolute pitch*: widely known as perfect pitch, which is a rare auditory phenomenon characterized by the ability of a person to identify or recreate a given musical note without a reference tone.

[4] *statistical regularities*: statistically regular patterns of syllables and statistically structured streams of sound created by speech segmentation. These regularities are different between languages (see François & Schön, 2014).

[5] *time-span reduction rule* (Jackendoff & Lerdahl, 1983, p. 146): the first rule of time-span segmentation (Jackendoff & Lerdahl, 1983, p. 146) is that "every group in a piece is a time-span in the time-span segmentation of a piece." The second rule of time-span segmentation is as follows: In the underlying grouping structure,

1. each beat $B$ of the smallest metrical level determines a time-span $T_B$ extending from $B$ up to but not including the next beat of the smallest level,

2. each beat $B$ of metrical level $L_i$ determines a regular time-span $T_B$, which is the union or sum of the time-spans of all beats of level $L_{i-1}$ (the next smallest level) from $B$ up to but not including

(a) the next beat $B'$ of level $L_i$ or

(b) a group boundary, whichever comes sooner, and

3. if a group boundary G intervenes between *B and the preceding beat of the same level, B* determines an *augmented time-span* $T'_B$, which is the interval from *G* to the end of the regular time-span $T_B$ (p. 147).

[6]The *foot* is the basic metrical unit that generates a line of verse in most Western traditions of poetry, including English accentual-syllabic verse and the quantitative meter of classical ancient Greek and Latin poetry. The unit is composed of syllables, the number of which is limited, with a few variations, by the sound pattern the foot represents. Contrasting with stress-timed languages such as English, in syllable-timed languages such as French, a *foot* is a single syllable.

The English word "foot" is a translation of the Latin term *pes*, plural *pedes*; the equivalent term in Greek, sometimes used in English as well, is *metron*, plural *metra*, which means "measure." The foot might be compared to a measure in musical notation. The foot is a purely metrical unit; there is no inherent relation to a word or phrase as a unit of meaning or syntax, although the interplay among these is an aspect of the individual poet's skill and artistry.

[7] In Anderson-Hsieh et al. (1992, p. 546), in "Figure 2. Standardized multiple regression coefficients for prosody, segmental error rate, and syllable structure error rate (N =60)," the high coefficient with prosody is mistakenly described as ($\beta$ = .058, $p$ < .001*)*, the second highest coefficient with segmental error rate ($\beta$ = -.027, $p$ < .001), and the third highest coefficient with syllable structure error rate ($\beta$ = -.025, $p$ < .001).

[8] *normalized pairwise variability index* (nPVI): Low and Grabe (2002) introduced the nPVI as a rhythmic index formulated as follows:

$$\text{nPVI} = 100 \left[ \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| /(m-1) \right]$$

*Where m = number of vowels in utterance*

*d = duration of the kth vowel*

[9]*phoneme deletion tasks*: Turkish and American children performed various phoneme deletion tasks with words in their native languages and pseudo-words. For the task, 20 common two- to five-letter words and 20 matched pseudo-words were used. The examples of verbal stimuli used in the experiments are follows:

The verbal stimuli for Turkish children (closest English translations)

| | words | pseudo-words |
|---|---|---|
| Initial vowel | ara (look for) | uca |
| | araba (car) | uvara |
| | ilik (warm) | itey |
| Initial consonant | tepsi (tray) | rekif |
| | davul (drum) | sakay |
| | fare (mouse) | folu |
| Final vowel | masa (table) | yire |
| | acele (hurried) | uyata |
| | idare (manage) | asata |
| Final consonant | pay (share) | piz |
| | ö rnek (example) | oktun |
| | temiz (clean) | pürün |

The vowel stimuli for American children

| | words | pseudo-words |
|---|---|---|

| | | |
|---|---|---|
| Initial word | edit | edel |
| | ocean | alam |
| | easy | opef |
| Initial consonant | dig | lim |
| | seven | puv |
| | very | sen |
| Final vowel | no | voo |
| | pie | nuy |
| | joy | edo |
| Final consonant | music | ufic |
| | us | ot |
| | sell | kallat |

All participants were tested individually in a private room at their own pace. They were taken through some examples, which showed them what to delete (the initial or final phoneme/note). No other instructions were given until the second block of 20 in each task, at which time participants were taken through some further examples, showing what to delete in the block. For the verbal items, the participants knew when the following items would be pseudo-words but not when the to-be-deleted phonemes would be vowels or consonants. After the first task, they were given instructions appropriate to the second task and examples just as in the first task. Participants were told to say each verbal item aloud, minus the critical phoneme. All responses were recorded on a portable tape-recorder with a built-in microphone, to be transcribed and scored later (Peynircioğlu, 2002, p. 71).

[10] *tone glide identification task*: tests tone glide identification with low-dipping, high-falling, mid-rising, and high-level, between 250 to 350 Hz.

[11] *Mandarin tone identification task*: In this task, the participants chose one tone from four typical Mandarin tones, which are high-level, rising, falling-rising, and high-falling. In the Mandarin intact identification task, listeners identified the target syllable when the entire syllable was presented. In the Mandarin silent-center identification task, listeners identified the tone of target syllables when only the first and final syllables of the word were pronounced.

[12] *input hypothesis*: This states that learners progress in their knowledge of the language when they comprehend language input that is slightly more advanced than their current level. Krashen called this level of input “i+1,” where “i” is the language input and “+1” is the next stage of language acquisition (Krashen, 1977).

For example, “material that is familiar to student (i) and a certain amount of unfamiliar material whose meaning can be induced from the content (1). When students are presented with input at the i + l level, they make use of ‘key vocabulary’ items (nouns, verb, adjectives, and sometimes adverbs" (Krashen, Terrell, Ehrman, & Herzog, 1984, p. 266).

However, "according to this explanation, comprehensible input is defined in terms of content word vocabulary; the contribution that knowledge of syntax can make to the process of understanding language is not taken into account" (Call, 1985, p. 777).

[13] *miniature language*: S → AP+CP+(CP)

S1 →{(bana luret)(tesi rolos)(suni)}

AP1→(bana luret)

CP1→(tesi rolos); CP2→(soni)

[14] *rhythm discrimination task*: The three probes and the 12 test items in this rhythm discrimination task were developed for this study. Each item consisted of two rhythmic patterns that were either identical or different. The items were presented with audiotapes. The time schedule for each item was as follows:

"Listen carefully"---1.5s---1st sequence (6s) ---"and" (2s) --- 2nd sequence (6s)

The rhythmic patterns were played at a constant pitch on a flute at an average speed of 80 quarter notes per minute, and were 6s long. The two patterns to be compared were separated by a 2s interval in which an "and" was spoken. The children had to decide which of the pairs were identical and which were different. Half were identical. The task was divided into three five-item sections, and the number of correct answers was assessed.

[15] the Fundamental Difference Hypothesis: Children are known to learn language almost

completely through implicit domain-specific mechanisms. Adults have largely lost the ability to learn a language without reflecting on its structure and have to use alternative mechanisms, drawing in particular on their problem-solving capacities, to learn a second language. (DeKeyser, 2000, p. 499)

[16] P1, N1, and P2 components: the P1 component reflects activity in the middle layers of the primary auditory cortex. The N1 and P2 components reflect activation of secondary auditory cortical areas.

Chapter 4 Study 2: Relationship between Sound Sensitivity, Receptive/Productive Prosody Processing of English, and English Listening Comprehension of Elementary School Children

[17] criteria to evaluate the Prosody production test: The evaluators came to a consensus on the criteria as follows:

G is marked for a good response (exact imitation) and counted as 1.0 point.

F is marked for a fair response and counted as 0.5 point: not exact repetition – stress/chunking exaggerated or minimized compared with stimulus – but function maintained, i.e., stress on the same word as in the stimulus, or information chunked as in stimulus.

P is marked for poor/incorrect response and counted as 0 point (e.g., stress on wrong word, chunking misleading).

[18] criteria to evaluate the Rhythm production test: The evaluators came to a consensus on the criteria as follows:

G is marked for an exact imitation of the rhythm sequence and counted as 1.0 point.

F is marked for a fair imitation and counted as 0.5 point. Accent minimized compared with the stimulus, but rhythm maintained as in the stimulus.

P is marked for poor/incorrect rhythm and counted as 0 point.

**Glossary** (in alphabetical order)

*B*: Unstandardized coefficient

$\beta$ (beta): Standardized coefficients, $-1 < \beta < 1$

Critical period: The concept whereby children have an advantage in ultimate attainment, not in rate of learning (Krashen et al., 1979; Slavoff & Johnson, 1995).

*Df*: Degree of freedom

Intercept: (x, 0), (0, y)

Mandarin Chinese tones and diacritics:

| Tone No. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Example for diacritics: | ā | á | ă | á | A |
| Example for the meaning of "ma" depending on the tone: | mother | fiber/hemp | horse | scold | (no meaning. like Japanese “ka,” a question marker) |

Meter: Implies measuring. It provides the means of such measurement for music; its function is to mark off the musical flow into equal time-spans (Jackendoff & Lerdahl, 1983).

Metrical feet: The principles for segmentation into feet vary among languages (Jackendoff & Lerdahl, 1983). In addition to monosyllabic feet seen in French, Vergnaud and Halle (1979) found three kinds of metrical foot-formation rules: those establishing bisyllabic feet, those establishing trisyllabic feet, and those establishing feet of unlimited length. English appears to have monosyllabic, bisyllabic, and trisyllabic feet, based on the number of syllables constituting a foot.

Metrical structure: The intuition that the events of the piece are related to a regular alternation of strong and weak beats at a number of hierarchical levels (Jackendoff & Lerdahl, 1983).

N: Sample size

*p:* Probability

Phonetics: the scientific study of speech sounds, including the study of the acoustic structure of speech and the mechanisms by which speech is produced and perceived (acoustic, articulatory, and auditory phonetics) (Patel, 2008, p. 37).

Phonology: the study of the sound patterns of language, including the study of how speech sounds are organized into higher level units such as syllables and words, how sounds vary as a function of context, and how knowledge of the sound patters of language is represented in the mind of a speaker or listener (Patel, 2008, p. 37). The science of speech sounds including, especially, the history and theory of sound changes in a language or in two or more related languages.

Prosody: The three aspects of word prosody are recognized, including stress, tone, and lexical pitch accent (Patel, 2008). These three aspects of word prosody are not mutually exclusive: There are tone languages with stress such as Mandarin and without it such as Cantonese, and pitch-accent language with stress such as Swedish and without it such as Japanese (Jun, 2012). A general term encompassing intonation, rhythm, tempo, loudness, and pauses, as these act together with syntax, lexical meaning, and segmental phonology in spoken text (Wennerstrom, 2001, p. 4).

*r*: Reliability coefficient. $-1 < r < 1$. $R < 0.20$ means no correlation, $0.2 < r < 0.4$ means weak correlation, $0.4 < r < 0.7$ means middle correlation, and $0.7 < r$ means strong correlation between the variables (Takeuchi & Mizumoto, 2012).

$R^2$: Squared multiple correlation coefficient of determination. $R^2 = .01$ means small accountability, $R^2 = .09$ means middle accountability, and $R^2 = .25$ means large accountability for the dependent variables (Takeuchi & Mizumoto, 2012).

Rote learning: quick learning without deep thinking, deduction, not induction

*SEB* = Standard error of the estimates (unstandardized coefficient)

Sound sensitivity: Throughout this thesis, sound sensitivity describes both or either of musical tone/rhythm perception skill and musical tone/rhythm production skill.

*t*-test = This is called Student's *t*-test, and is one of the statistical hypothesis tests.