교육학석사학위논문

Korean High School Learners' English Rhythm and Comprehensibility

한국 고등학교 학습자들의 영어리듬과 이해도

2015년 8월

서울대학교 대학원

외국어교육과 영어전공

문 예 지

Korean High School Learners' English Rhythm and Comprehensibility

by
Yeji Mun

A Thesis Submitted to
the Department of Foreign Language Education
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Education

At the Graduate School of Seoul National University

August 2015

ABSTRACT

Since English has attained an important role as an international language, a number of diverse global dialects have become recognized. This has had a significant impact on the area of English pronunciation acquisition as the goal to be achieved is being changing with the major focus now shifting from phonemic level targets to the overall intelligibility. Understanding the suprasegmental rules appears central to this new goal, with particular mention given to stress and rhythm as the important elements. Yet, Korean learners are subject to make inappropriate rhythm units opposed to the nature of English rhythm, in that Korean belongs to syllable timing and English to stress-timing. Therefore, the present study proposed the three research questions as follows: (a) are there differences in speech rhythm between the Korean EFL learners and the native speakers of English?; (b) are there differences in speech rhythm between the more proficient and the less proficient learners?; (c) does Korean EFL learners' sentence stress affect native speakers' comprehensibility of speech rhythm?

The two main experiments to solve the research questions were implemented with a total of 21 Korean EFL high school students and 3 native speakers. The first experiment was divided into the two sub-tests: the one was to investigate the speech rhythm comparing the score of nPVI-V and nFPVI between Korean learners and native speakers, and the other was to apply the

same metrics measuring two 6 more proficient groups against 15 less proficient groups to manifest the different effects based on the proficiency level. The comprehensibility test was conducted by 3 native speakers in order to assess how much the learners' speech rhythm affects the native speakers' comprehensibility.

The major findings of the present study suggest that the Korean learners did not seem to use a stress pattern to achieve the English rhythm. There were significant different results of the nPVI-V scores between the non-native and the native speakers. One argument yet to be discussed was the fact that nFPVI did not seem to support the typology of two different languages in terms of the rhythm unit. With regard to the nature of the regular occurrence of stressed and unstressed syllables in English, a foot unit is supposed to generate an English isochrony. In this respect, the native speakers were expected to have the lower scores of nFPVI and the Korean speakers were anticipated to make the higher scores of nFPVI. Nonetheless, the results failed to show the different nFPVI between the two language groups and the two groups of the different proficiency level. Furthermore, the scores of comprehensibility ratings based on the speech rhythm serve to indicate that the speech rhythm might affect the native speakers' comprehensibility. Therefore, it can be concluded that Korean learners of English do not appear to make the English speech rhythm when conducting speech performance. These findings present the understandings on the speech rhythm by Korean EFL learners, and contribute to illuminating the importance of stress and

the rhythm unit on the pronunciation teaching.

Key Words: Prosody, Speech Rhythm, Stress-timing, Syllable-timing

Student Number: 2011-21524

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The primary aim of the present research is to investigate speech rhythm of Korean EFL learners and the native speakers' comprehensibility. This chapter consists of three sections: Section 1.1 presents the purpose of the study to provide overall framework for the reason why the present research is needed. Section 1.2 develops main research questions in terms of the first section. Finally, Section 1.3 offers a general outline of the organizations of the thesis.

1.1 Purpose of the Study

With English now taking the role of global lingua franca (English as an international language), there have been calls to revise current understandings of acceptability that incorporate speakers' wider range of accents in different L1 backgrounds (Jenkins, 2004). Many now recognize the variety of L1 influenced English accents as having legitimacy and, as such, intelligibility rather than accuracy of speech is a new emphasis (Crystal, 1996; Saito, 2012; Breitkreutz, Derwing, & Rossiter, 2001; Derwing & Munro, 1995, 1997; Gilakjani, 2012;

Celce-Murcia, *et al.*, 2010; Caspers, 2010). This has particular bearing in the area of pronunciation instruction. In terms of pedagogy, a primary concern is understanding and uncovering phonological rules that are aimed at improving communicative competence.

Opposing the traditional pedagogical approach in English language education in Korea which focused on achieving the native-like pronunciation according to L1 models of standard American English, 2011 guidelines of Ministry of Education and Science Technology detail the inclusion of suprasegmental development, in addition to phonemic level practice, as primary goals. They specifically note the importance of the sound and meaning relationship and list facets such as connected speech, prosodic changes of speech rate and natural production and prosodic changes in a certain situation. Table 1.1 presents application of these according to listening and speaking objectives in the 2009 Revised National Curriculum (2011):

Table 1.1
Objective for Listening and Speaking

Listening	Speaking
✓ Students can understand various types of English rhythm, stress, and intonation.	✓ Students can speak based on English rhythm, stress, intonation, and linked sound.

(2009 Revised National Curriculum, 2011, p. 23)

Given the recent shift and pedagogical objective, consequently, it can be addressed that learners are required to acquire an ability to make comprehensible pronunciation which will help to develop communicative abilities.

With a view mentioned above that there has been a transition from accurate pronunciation to intelligible communication acclaimed by many researchers (Celce-Murcia *et al.*, 2010; Wong, 1987; Derwing & Rossiter, 2003), despite a few clear answers of which specific components in suprasegmentals should be taught, some researches have supported the idea that one of the most important elements among prosodic features to be learned is stress pattern.

Celce-Murcia *et al.* (2010) insisted that connected speech was closely related to the suprasegmental features to be better understood in communication, for a wide range of meaning was delivered through the suprasegmental property than segmental features. To be specific, stress and rhythm structure were more

connected to misunderstanding when they were poorly performed, while intonational contours were more linked to politeness as well as awkwardness when inappropriately used.

In light of the mutual relation between the stress patterns and a better intercommunication, Field (2005) investigated the lexical stress as one factor that language learners were needed to acquire. It was supported that a major function of lexical stress was to grant the tendency that native speakers were more likely to depend on stress syllables to detect the word and the intended meaning. The results attested that stress shift in content words could directly affect an overall understanding, suggesting that it might be desirable to teach and learn lexical stress for a better intelligibility.

In the area of language process and perception, other supportive findings lay in prosodic sensitivity of newborns who were able to distinguish languages of different rhythm units. Nazzi, *et al.* (1998) provided infants with the manipulated sentences in which segmental variables were eliminated other than prosodic cues. The result was that the French babies had an ability to differentiate stress-timed and syllable-timed languages, whereas they did not distinguish between English and Dutch which are categorized as the same rhythm class.

Such evidence corresponded to Ramus (2002), where intonational cues

in each stimulus were artificially modified to see how French newborns perform language discrimination task using rhythm property. The study suggested that infants seemed to know how to apply the rhythmic cues of their native language when identifying other languages regardless of intonation, thereby they could differentiate between Dutch and Japanese which belong to the different rhythm categories.

All taken into account, prosodic features can play a key role to understand and deliver meaning for better communication. In the present study, consequently, the English speech rhythm will be investigated between native English speakers and Korean EFL high school learners based on the rhythm measures and comprehensibility rating by native speakers.

1.2 Research Questions

The present study seeks to examine the realization of speech rhythm of Korean EFL high school learners by rhythm measures and illuminate the relationship between speech rhythm and comprehensibility. In the expectation of observing the perceived rhythm (e.g. comprehensibility) as well as an acoustic quantification of the speech rhythm, there emerge the following research questions:

(1) Are there differences in speech rhythm between the Korean EFL high school learners and the native speakers of English?

(2) Are there differences in speech rhythm between the more proficient and the less proficient learners in Korea?

(3) Does Korean EFL high school learners' sentence stress affect native speakers' comprehensibility of speech rhythm?

1.3 Organization of the Thesis

The present thesis comprises 5 chapters. The purpose of the study and the main research questions are introduced in Chapter 1. Chapter 2 expatiates the theoretical background and empirical research of speech rhythm and comprehensibility. Chapter 3 elaborates the methodology adopted in the present research and introduces the procedure of the experiment. Chapter 4 interprets the results of the experiments and discuss the conclusions. Chapter 5 encapsulates the major findings and suggests the pedagogical implications as well as the limitations for the further research.

CHAPTER 2
LITERATURE REVIEW

This chapter presents a general overview of the previous studies on the speech rhythm, rhythm measure, and the comprehensibility with regard to the research questions presented in the previous chapter. Section 2.1 describes the nature of sentence stress and speech rhythm in connected speech. Section 2.2 introduces rhythm metrics to distinguish speech rhythm based on the conventional rhythm typology. Section 2.3 provides evidence of how speech rhythm is significantly related to comprehensibility.

2.1 Speech Rhythm

Rearranging the nature of syllables in connected speech is one of the most conspicuous traits of rhythm patterns in English. What gives a detailed account of this phenomenon is regular alternation between stressed and unstressed syllables. Even though stressed syllables are conventionally defined as longer, louder, and higher in pitch (Ladefoged, 2006), it can be differently explained according to the point of view between speakers and listeners (Celce-Murcia *et al.*, 2010).

For instance, stress can be regarded as an outburst of energy made of air pressure which is pushed from the lungs so that the speaker is able to produce stressed syllables. When the listener hear the utterances from the speaker, stress can be considered as one of the distinct features which contains slightly longer vowel durations and higher pitch in the stressed syllables. In terms of the listener's point of view, stressed syllables should be discriminated from unstressed syllables in terms of the duration of syllables. This aspect is closely related to the complexity of onsets and codas of syllables, and full vowels can appear intermittently in accordance with the placement of stress. As a result, those who have a native language in which durational differences are not significant between stressed and unstressed syllables (e.g. syllable-timing) might need to develop different stress patterns.

As Field (2005) mentioned lexical stress was important in order not to breakdown an intelligible communication, the sentence stress is also critical to understand the overall meaning in connected speech. For the purpose of creating the rhythmic beat in English utterances, each sentence can rearrange the stress location based on the placement of content words. When content words and function words are placed in sentences or speech, stressed syllables tend to be located on content words which deliver major information except for function words which convey grammatical meanings. Words are categorized as content

words and function words in Table 2.1.

Table 2.1
Content Words and Function Words

Content Words	Function Words
nouns	articles
main verbs	auxiliary verbs
adjectives	personal pronouns
possessive pronouns	possessive adjectives
demonstrative pronouns	demonstrative adjectives
interrogatives	prepositions
not/ negative contractions	conjunctions
adverbs	
adverbial particles	

(Celce-Murcia, 2010, p.212)

This trend can be easily witnessed in English rhythm drills presented in Figure 2.1. Notwithstanding the different numbers of syllables in all the sentences, the number of stressed syllables is identical, which results in the same time span for each sentence. If the sentence has more unstressed counterparts, the reduction of more syllables among stressed syllables arises. In this regard, it can

be assumed that the rhythm system in English is highly dependent on the occurrence of regular intervals of stress as much as the natural speech flow is recognized.

Figure 2.1
'Cats chase mice' rhythmic drills

	CATS		CHASE		MICE.
The	CATS	will	CHASE		MICE.
The	CATS	have	CHASED	the	MICE.
The	CATS	have been	CHASing	the	MICE.
The	CATS	might have been	CHASing	the	MICE.

(Celce-Murcia, 2010, p. 210)

Great attention, subsequently, has been drawn to the question of what speech rhythm is in general and a lot of studies have attempted to explore isochronous units in connected speech to define rhythm itself. In this fashion, most of studies explained that rhythm structure is strictly related to the timing when each syllable in sentences is placed under particular periodic patterns, especially when speech has smooth flows (Ashby & Maidment, 2005; Steel, 1779; Pike, 1945; Lloyd James, 1940; Abercrombie, 1967). For instance, Abercrombie (1967) elaborated "rhythm emerges clearly during those moments

when speech is fluent and uninterrupted." In the similar breath, Pike (1945) described "a sentence or part of a sentence spoken with a single rush of syllables uninterrupted by a pause" is a rhythm.

Centering on syllables and syllable length for the tendency of the circular occurrence of rhythm, Steel (1779) manifested that stressed syllables tend to be placed at regular intervals in connected speech of English, which contributes to a shorter duration of non-stressed syllables to yield almost the uniform occurrence of feet. The regular beat, therefore, produced by rearranging the duration of stressed and non-stressed syllables in connected speech is related to the time equality, producing an English isochrony. In this point, stress and foot units can be the important rhythmic factors in English.

However, the units of English isochrony such as stress and foot cannot embrace all types of language rhythm, so that certain issues about speech rhythm distinction have claimed more attention. Conventionally, most of researchers have asserted that each language belongs to a particular prosody group generating different speech rhythm. Pike (1945), at the very first, introduced such distinction by dichotomizing the type of rhythm, "stress-timed" and "syllable-timed." Based on this typology, Abercrombie (1967) contributed to classifying all languages around the world as follows:

> As far as is known, every language in the world is spoken with one kind of rhythm or with the other. In the one kind, known as a *syllable-timed* rhythm, the periodic recurrence of movement is supplied by the syllable-producing process: the chest-pulses, and hence the syllables, recur at equal intervals of time – they are isochronous. French, Telugu, Yoruba illustrate this mode of co-ordinating the two pulse systems: they are syllable-timed languages. In the other kind, known as a stress-timed rhythm, the periodic recurrence of movement is supplied by the stress-producing process: the stress-pulses, and hence the stressed syllables, are isochronous. English, Russian, and Arabic illustrate this other mode: they are stress-timed languages.
>
> (Abercrombie, 1967, p. 97)

Besides, this standpoint of the speech rhythm was adopted by Dauer (1983), giving a birth of the more sophisticated classification of languages based on the rhythm typology in Table 2.2.

Table 2.2
Category of stress-timed and syllable-timed languages

Stress-timed languages	Syllable-times languages
English	French
Russian	Spanish
Germanic	Yoruba
Arabic	Telugu
Thai	Hindi
Brazilian Portuguese	Indonesian

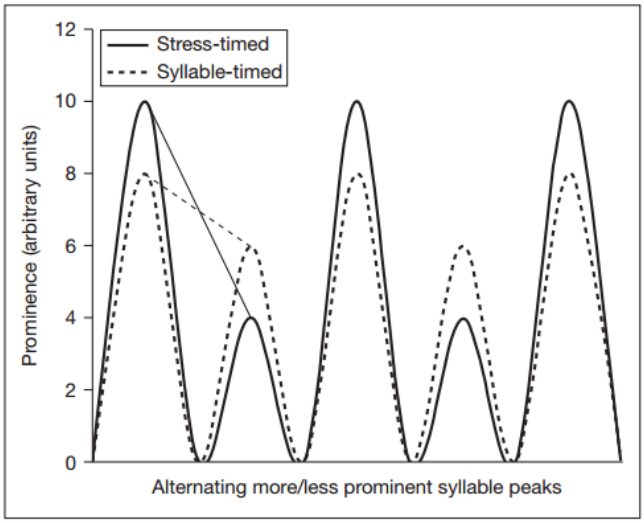
(Partially excerpted from Dauer, 1983, p. 56)

In this point of view, as English is stress-timed and Korean is syllable-based language (Ramus, 2002; Abercrombie, 1976; Pike, 1945; Mock & Lee, 2008), there are supposedly different ways of making speech rhythm between the two different types of languages. In stress-timed languages, as mentioned above, stress is regarded as an element of English isochrony by making certain syllables and words prominent with longer, louder, or higher pitch production (Lee, 1997; Han, 2015; Ladefoged, 2006). When placed at regular intervals, the stress patterns lead to a successive occurrence of both stressed and unstressed syllables in turn, as well as bring a rhythmical stream, foot, in phrases, sentences and connected speech (Han, 2015; Dalton & Seidlhofer, 1994; Patel *et al*., 2006). On

the other hand, in syllable-timed languages each syllable does not meet this condition to make isochronous units. Regardless of the stress patterns there are almost the same length of syllable durations generated in speech.

As Figure 2.2 shows the prominence gradient between stress-timed and syllable-timed languages, the prominent syllable peaks of stress-timed languages are higher than those of syllable-timed languages. Furthermore, the height of the peak in the stressed-timed languages is more equivalent than that of the syllable-timed languages. The gap between stressed and unstressed syllables in the stress-timed languages are more widened than that in the syllable-timed languages. With regard to the function of stress as one of the distinct characteristics of English, it might be recommended for Korean learners who acquired the syllable-timed language as a mother tongue in order to attain the stress patterns and know how to produce English rhythm for better communication.

Figure 2.2
Schematic representation of the prominence gradient



(Asu and Nolan, 2009, p. 67)

2.2 Measurement of Speech Rhythm

The rhythm measures have been investigated to quantify speech rhythm in languages to demonstrate the traditional rhythm typology: stress-timing and syllable-timing. In order to understand the main flow of the rhythm classification, the present research mainly concentrated on providing the rationales of the rhythm measures as follows.

2.2.1Rhythm Measurement Methods

Many empirical studies have endeavored to support the rhythm class, proposing the various types of rhythm metrics. Jang (2009) introduced the five instrumental formulas which were devised to classify speech rhythm categories and the other three metrics formulated to differentiate between native and non-native speakers. The five rhythm metrics mentioned in Jang's research included %V, Varco C, Varco V, nPVI-V, and rPVI-C.

%V was originally introduced by Ramus *et al*. (1999) including ΔV and ΔC, all of which were designed for measuring durations of the vowel and consonantal intervals. The definition of each formula is provided in Table 2.3.

Table 2.3
Definition of Formulae

Formula	Definition
%V	the proportion of vocalic intervals in the sentence
ΔV	the average standard deviations of vocalic intervals within the sentence
ΔC	the average standard deviations of consonantal intervals within the sentence

(Ramus, 1999, p. 7)

%V and ΔV metrics were designed to calculate vowel intervals, i.e., vocalic intervals, while ΔC employed consonantal intervals, i.e., intervocalic intervals. Table 2.4 shows the definition of the two intervals.

Table 2.4
Definition of Vocalic and Intervocalic Intervals

Type	Definition
Vocalic Intervals	the duration of sequences of consecutive vowels (from onset of the first vowel to offset of the last vowel of the sequence)
Intervocalic Intervals	the duration of sequences of consecutive consonants

(Ramus, 1999, p. 7)

According to the first acoustic correlate of speech rhythm, %V is related to the continuous time that can be elongated when the speaker whose native language is syllable-timed articulates the target language which is stress-timed. This is because syllable-timed languages have less complicated syllable structure and are supposed to yield higher figure than stress-timed languages which have a complex fluctuation of syllables. In contrast, ΔV and ΔC are expected to have a higher value in stress-timed languages than in syllable-timed languages, for the vocalic intervals can vary due to the nature of the fluttering prosodic features in

the speech sound of stress-timed languages.

One of the fatal defects of the three metrics, however, was the fact that the scores were extremely sensitive to speech rate, making it less effective to account for cross-linguistic comparison. In order to compensate the defect, White and Mattys (2007), and Dellwo and Wagner (2003) designed Varco V and Varco C, respectively. The fundamental difference between ΔV and ΔC, and Varco V and Varco C was that the latter two metrics normalized the speech rate, which accordingly less affected the results of the rhythm measurement.

Varco V is calculated as the standard deviation of vocalic interval durations is divided by a mean vocalic interval duration and multiplied by 100. Varco C was devised to control the consonant durations as for the view that there would be longer consonantal intervals in a slow utterance. Varco C normalizes the speech rate as the standard deviation of intervocalic interval durations is divided by a mean consonantal interval duration and multiplied by 100. Each formula is presented as follows:

$$\text{Varco V} = \frac{\Delta V \times 100}{mean\ V} \quad \text{(V = duration of vocalic intervals)}$$

$$\text{Varco C} = \frac{\Delta C \times 100}{mean\ C} \quad \text{(C = duration of consonantal intervals)}$$

In parallel with the approach to controlling the speech rate, but with one step taken forward, Low and Grabe (2002) introduced a normalized Pairwise Variability Index (nPVI-V) as well as a raw Pairwise Variability Index (rPVI-C) to shed lights on the durational differences of successive vowel and consonantal intervals. A noteworthy trait was that the Pairwise Variability Index (PVI)[1] focused on the durational difference between adjacent intervals instead of the total length of all intervals regardless of measuring either vocalic or intervocalic intervals. It was in virtue of the prediction that the successive intervals would have a huge duration gap due to a regular alternation of stressed and unstressed syllables in stress-timed languages. What differed from the rPVI-C to the nPVI-V was a matter of normalization. The former was administered to calculate the mean value of the difference between adjacent intervocalic intervals, which can be variable and language specific depending on the syllable-structure (Asu & Nolan, 2005). For example, in English phonotactics, the maximum number of consonants before and after a vowel is three, whereas the number of consonants

[1] PVI mainly referred to rPVI-C in the previous research. However, in the present research, the two terms will be differently used in a way that PVI means a variety of the PVI formulae regardless of normalization or the type of intervals. Meanwhile, rPVI-C will represent the PVI formula which is applied intervocalic intervals only.

before and after a vowel is one in Korean phonotactics. Although the number of consonants affects the rPVI-C value and differs between stress-timed and syllable-timed languages, the different number of consonant arrangements can be accepted as the nature of languages to be compared. Low and Grabe (2002), for this reason, suggested that the difference of consonantal durations should be taken into account without normalizing the speech rate. The raw PVI for intervocalic intervals simply numerates the mean of the difference between consecutive intervals:

$$\text{rPVIC} = 100 \times \left[\sum_{k=1}^{m-1} \frac{|d_k - d_{k+1}|}{m-1} \right]$$

In the rPVI-C formula, *m* is number of intervocalic intervals and *d* is the duration of the *k*th interval. The results of the rPVI-C value are assumed to be lower when there are few dynamic differences in consonantal intervals. That is, the more they differ, the higher the index is supposed to be. In the same line with this conjecture, Korean learners of English are anticipated to have a lower value than native speakers as Korean has one consonant according to Korean phonotactics.

In contrast to the rPVI-C, the nPVI-V adjusts the speech rate for

normalization of the result value, dividing each difference of adjacent vocalic intervals by the average duration of two successive syllables in pair. In this fashion, the possibility that the results of speech rhythm across speakers are affected by a speech tempo can be possibly reduced. The following formulates the nPVI-V:

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

m is number of vocalic intervals in an utterance and *d* is the duration of the *k*th interval. The difference in duration between each pair of successive intervals is calculated, the absolute value of the difference divided by the mean duration of each pair is taken. The differences are all summed and then divided by the number of differences. The final output is multiplied by 100, in order not to yield the fractional values. It is anticipated that Korean speakers of English might have a lower value than native speakers like the rPVI-C. Nevertheless, the homogeneity of the result between the rPVI-C and the nPVI-V needs to be explained in a different way. The former value results from a great discord in duration of the consonantal intervals, whereas the latter value arises from the duration gap between the two consecutive vowels due to alternation of sentence

stresses.

As to rhythm measurement tools, most of the extensively utilized metrics to the rhythm research in connected speech including the rPVI-C and the nPVI-V are still in a controversial position on account of a lack of the empirical evidence to buttress the consistent and crystal-clear outcome. There might remain inexhaustible acoustic or prosodic aspects the previous research did not recognize yet. For example, Wagner (2014) inspected whether Polish speech rhythm belonged to either stress-timing or syllable-timing with %V, ΔV, ΔC, rPVI-C, Varco C, and Varco V. Not all the scores that Polish attained were dependable because the low scores of Varco V and nPVI-V could support Polish as the syllable-timing, the obscure scores of Varco C and % V indicated the intermediate position between the two categories, and the high scores of ΔC and rPVI-C signified Polish was the stress-timing. The Euclidean distances implemented was also in the same line with the previous findings. The % V-nPVI-V distance associated Polish with syllable-timed languages (e.g. Italian and Spanish), while the Varco V-Varco C distance related Polish with stress-timed languages (e.g. German).

There have been calls for unearthing some factors contributing to the unsuccessful consequences of rhythm research, therefore, and the multifaceted approaches for the speakers to satisfy the target languages were, albeit

marginally, uncovered: the phrase-final lengthening and the elongated articulation of the stressed syllables produced by the syllable-timing speakers seemed to result in the relative shortening of the neighboring weak syllables. It was also presumed that this phenomenon led the speech rhythm articulated by language learners to be marked in the ambiguous boundary between the syllable- and the stress-timing (Wagner, 2014).

Arvanit (2009) also stated that there were several reasons why syllable-timed languages made a high PVI score. For example, Korean and Spanish speakers of English both achieved as a high score of rPVI-C as native speakers. The given possibility was that Korean speakers made a phrase-final lengthening and Spanish speakers produced consonant clusters with a weak forced pronunciation. It was assumed that the two different tactics to process the given speech caused the intervocalic intervals had a similar high score like native English speakers. The findings from the two research mentioned above suggested that not only did participants use the different strategies which appeared to affect the native-like scores of the rhythm metrics, but also the rhythm measures did not clearly discriminate a certain language based on the dichotomized rhythm typology. Notwithstanding the major findings that would be worthwhile understanding the comprehensive nature of the speech rhythm, more generalization would be needed with a wide scope of corpora from various

types of languages (Ramus, 2002).

In spite of incessant struggle to succeed in categorizing speech rhythm, one aspect overlooked by the previous research which mainly concentrated on vowels and consonants was the fact that English isochrony (i.e., foot) occurs in alternation of stress and unstressed syllables. Even though most researchers agreed that stress-timed languages consist of rhythmic beat based on regular appearance of feet, there have been a lack of examining rhythm measurement by acoustic correlates based on feet. Ramus (2002) proposed that it might be feasible to substitute the foot for rhythmic variables – such as vowels and consonants – as an alternative unit for measuring speech rhythm. It was assumed that the feet have been ignored on account of the belief that rhythmic qualities could be entirely exploited by the syllable features. In order to materialize a rhythm metric with feet, Asu and Nolan (2005) devised a normalized foot PVI (nFPVI).

Traditionally, the feet have been defined as a unit that bears only one stressed syllable and more than one consecutive unstressed syllable. In more details with phonetic terms, the feet are a unit which ranges from the onset of a stressed syllable to right before the beginning point of the onset of the next stressed syllable (Giegerich, 2005; Ashby & Maidment, 2005; Celce-Murcia *at al*., 2010; Ladefoged, 2006).

Although Asu and Nolan (2005) accepted the conventional rhythmic stress unit and calculated foot durations, Asu and Nolan (2006) modified the previous norm of the feet in that unreduced vowels could also consist of a foot. In order to identify the feet unit, it was assumed that 'full vowel strategy' would be applicable due to the difficulty in detecting rhythmic stresses as well as the possibility that unstressed vowels sometimes are pronounced as the full vowel. For example, 'modest gymnasts' [mɒdəstdʒɪmnæsts] have the three feet ranging from [ɒ] to [ə], from [ə] to [ɪ], and from [ɪ] to [æ].

There has been, however, little research applying the feet to nPVI and a lack of supporting the rationale of availability to conduct the full vowel strategy for detecting the feet. The present research will accordingly accompany the traditional notion of feet in an acoustic view with the nFPVI rhythm measure by Asu and Nolan (2005, 2006, 2009).

2.2.2 Necessity of Normalizing Speech Rate

While most of acoustic correlates mainly aimed to measure syllable durations, it still remained in question to attain the divergent results failing to control the speech rate (Grabe & Low, 2002; Ramus, 2002; Dellwo & Wagner, 2003; Dellwo, 2006; White & Mattys, 2007; Stockmal, *et al.*, 2005). With a view to normalize a tempo of speech segments White and Mattys (2007), and Dellwo (2006) introduced Varco V and Varco C, respectively.

The former study revealed that without control of speech rate, a) ΔV measure did not clearly distinguish rhythm categories as French belonged to the category of stress-timing such as Dutch and English, and b) ΔC and rPVI-C identified French and Dutch in the midway between English and Spanish. Varco V, on the other hand, as a normalized measure, was in parallel with the result of nPVI-V. The latter study disclosed that Varco C seemed to vary among rhythm classes in that the faster speech rate decreased the value of Varco C. To sum up, ΔC and rPVI-C were mostly vulnerable to the speech rate variable due to a lack of normalization (Stockmal *et al.*, 2005; Dellwo & Wagner, 2003).

To figure out the stable result to distinguish the language rhythm categories, most research acknowledged that PVI might be more probable to grasp the overall rhythm structure and detect the different realization between

stress-timing and syllable-timing (Gibbon & Gut, 2001; White & Mattys, 2007; Mok & Lee, 2008; Tan & Low, 2014; Grabe & Low, 2002). It can be supported by Dellwo *et al*. (2009) in the aspect that both nPVI-V and rPVI-C were best suitable to show the discrimination of native English speakers from Spanish Speakers of English. Although most rhythm metrics supported the rhythm typology, unsettled results due to some variables (e.g. speech rate) did not support all rhythmic measures to be used. Consistent output compared to other types of metrics appeared in the PVI formula. Gibbon and Gut (2001) stated that the PVI succeeded in distinguishing rhythmic differences between stress-timing and syllable-timing. When White and Mattys (2007), and Tan and Low (2014) applied Varco V and nPVI-V as rhythm measures, the former concluded that both metrics were more reliable than others. The latter suggested that the significant difference was found in nPVI-V not in Varco V. Mok and Lee (2008) implemented %V and nPVI-V which summarized both metrics were dependable to discriminate different rhythm patterns. Among scattered findings to select the better rhythm metrics from the previous research, nPVI-V in common showed a consistent and reliable output.

For this reason, the present research is intended to inspect speech rhythm with nPVI-V rather than rPVI-C so as to avoid the instability and to solely measure the acoustic value related to vowels and syllables which can be

standardized by the normalized metrics. Furthermore, nFPVI will be added to the present research to uphold the necessity of isochrony when producing speech rhythm. Other metrics (e.g. %V, Varco V, Varco C, and rPVI-C) will not be measured in the current study owing to the ambiguous results of the previous research.

2.3 Comprehensibility on Speech rhythm

It has been claimed that intelligibility and comprehensibility of listeners can be affected by speech production of speakers. (Gow & Gordon, 1993; Derwing & Munro, 1995, 1997; Winters & O'Brien, 2013; Frazier *et al*, 2006, Anderson-Hsieh & Koehler, 1988) According to the previous research, intelligibility refers to "the extent to how listeners understand the message in the speech produced by the speakers", while comprehensibility refers to "listeners' perception of how difficult it is to understand the speakers' utterances (Derwing & Munro, 1995, 1997)".

Among various factors which influence intelligibility and comprehensibility did some research elaborate the effects of prosody on perception process, ranging from word stress intelligibility task (Field, 2005) to

sentence stress perception task (Gow & Gordon, 1993), as to the aspect that prosody matters in connected speech especially when decoding a meaning (Cutler *et al.*, 1997; Tyler & Warren, 1987; Frazier et al, 2006; Winters & O'Brien, 2013).

Field (2005), for instance, explained the effects of stress on intelligibility. When the lexical stress was misplaced, both native and non-native subjects showed a similar result of misunderstanding the given words. Therefore, stress can influence intelligibility, non-native English learners might need to practice speech decoding through applying lexical stress patterns.

A further research including comprehension tests based on the prosody perception were carried by Han (2004). The native English speakers were requested to perform the recall task and the comprehension checkup quiz after listening to the discourse of the language learners. When the primary stress was well placed in given acoustic stimuli, the reaction time to choose the answer was faster and the test score was higher than the stress was poorly located. It can be suggested that the correct stress placement boosted the better comprehension. However, comprehension tests conducted in the experiment of Han (2004), in principle, were more like intelligibility task: one task was a recall test to write down information as much as possible, and the other was to answer the given questions based on the listening stimuli.

More developed studies of speech perception with prosodic cues were provided by the empirical research. Frazier *et al*., (2006) mentioned that as soon as listeners received the acoustic input from speakers, they automatically anticipated and used a series of features derived from the input under the process of sound analysis to decode the syntactic structure based on their native language. Furthermore, even when speakers were required to pronounce the ambiguous sentences to deliver one certain meaning of the two, they were able to produce appropriate utterances with adequate prosodic cues, so that listeners could pinpoint the meaning that speakers intended to say (Allbritton *et al.*, 1997). It seemed, therefore, to support the claim that prosody might well be one of the important factors to understand connected speech.

Winters and O'Brien (2013) investigated effects of segmental and suprasegmental elements on the perception of foreign accent and intelligibility in the first language (L1) and second language (L2). The research specifically centered on how intonation influences the speech perception in suprasegmental experiment and concluded that when intonation was acoustically manipulated the tokens sounded less intelligible to both the L1 and L2 listeners. It seemed that prosody might affect the perceptual processing of overall intelligibility in that non-native intonation reduced intelligibility rather than segmental cues.

The emphasis on the current study is that the primary inquiry to be

answered is the extent to how difficult it is to understand English rhythm of Korean speakers, now that the more difficult it is to understand the sentence stress, the more interruption there might be to grasp a good speech flow. The present study, thereupon, is to delve into accounting for the relationship between the perceptual difficulty (i.e., comprehensibility) and speech rhythm instead of diagnosing decoding process (i.e., intelligibility). In addition, it is a main issue to inquire whether the sentence stress affects the perceptual difficulty in understanding the speech rhythm. Therefore, the norm of comprehensibility here is to be limitedly used only for judging how difficult it is to understand speech rhythm based on the sentence stress.

CHAPTER 3
METHODOLOGY

This chapter provides the methodology applied for the present study. Section 3.1 describes participants, the research setting, and the materials of the study. Section 3.2 elaborates the dada coding procedures and demonstrates a data analysis.

3.1 Experimental Methods

3.1.1Participants and Setting

A total of 24 participants, including 21 Korean learners of English (NNS) and 3 native English speakers (NS), participated in the present research. All of Korean participants in this experiment were the first graders in high school located in Seoul who could speak a standard Seoul dialect. Table 3.1 shows the details of all the subjects' background information.

Table 3.1
Korean Participants (NNS)

	More proficient group (MP)	Less proficient group (LP)
L1	Korean	Korean
Gender	male 4, female 2	male 3, female 12
Mean Age	17	17
Mock English Scores **Mean test score (listening score)*	M= 84.3333 (33.1667)	M= 45.7333 (27.1333).

Mock English Listening test score was used in implementing an independent sample t-test to confirm that the two groups are different in terms of the total test score as well as the listening score. The results of an independent sample t-test for the total test score and the listening score in Table 3.2 shows that there is significant difference between more proficient group and less proficient group in test score ($p<.05$) and listening score ($p<.001$) respectively.

Table 3.2
Independent samples T-Test Summary of Test Score and Listening Score

	Mean		SD		95% CI for Mean Difference		t	Sig.
	MP (n=6)	LP (n=15)	MP	LP	Lower Bound	Upper Bound		
Test Score	33.1667	27.1333	3.48807	5.46243	-11.10741	-.95926	-2.489	.022*
Listening Score	84.3333	45.7333	6.47045	9.14539	-47.21722	-29.98278	-9.375	.000*

To compare English speech rhythm of Korean speakers to that of native speakers, 3 native speakers participated in the present research. As shown in Table 3.3 all of them are from the U.S.A., and their gender is the same as female. All the English speakers can speak American English.

Table 3.3
Native English Participants (NS)

	Native English Participants (NS)		
Participant #	1	2	3
L1	English	English	English
Gender	female	female	female
Age	20	20	36
Nationality (State)	USA (Connecticut)	USA (Texas)	USA (New York)

3.1.2 Materials and Procedures

The recording experiment of the present study was conducted in a quiet room. At first, each student was provided with an overall guidance about how to perform the task properly. After the explanation all the participants were told to practice twelve sentences written on a reading sheet.

All the words on the reading list mainly consisted of twelve sentences: nine sentences excerpted from Lee (2013) and three fillers from Kim (2015). On the purpose of bringing the uniformity in each sentence, simple declarative sentences were selected. Vocabularies comprising target sentences were also adjusted on the level that the first grade high school students could fully elicit the stored memory from their lexicon. The word list for the adaptation was referred to the English education section of the 2009 Revised National Curriculum (2011).

For the sake of speakers not to be involved in a possible interference in speech production because of the length of the sentence, all the nine sentences except for three fillers were modified based on the number of syllables; short, medium, and long. This manipulation of the target sentences might be meaningful in that most of the previous research barely controlled the target sentences to be used. The manipulated target sentences and the fillers can be seen in Appendix A and Appendix B, respectively.

NNS participants were required to read randomly listed twelve sentences

for the recording. Voice recording was conducted in a quiet room with PCM-D50 SONY, which is *96 kHz*/24-bit recorder accommodated with two-position stereo microphones. Students were given ten minutes to practice reading sentences and recorded their voice twice. If needed, the researcher taught students how to pronounce certain words in target sentences.

Lastly, both the NNS and NS participants were required to fill out the survey sheet after the recording session was finished in order to collect the comprehensive background information of each participant. As shown in Appendix F, the survey sheet for Korean learners mainly consisted of whether the students had pronunciation instruction before. At the end of the experiment, the NS subjects also filled out the survey sheet for the language background shown in Appendix G.

When the native speakers finished the recording, they were asked to rate the NNS speakers' rhythm comprehensibility based on the sentence stress by 7 point Likert scale. In the previous research, as for the rating of intelligibility, it was necessary to carry out transcription of the non-native speakers' speech recordings, whereas comprehensibility rating was conducted by marking the Likert-scale. 9 point Likert scale has been mostly used to distinguish the level of perception difficulty in non-native speakers' speech: from 1 (extremely easy to understand) to 9 (extremely difficult or impossible to understand) (Derwing &

Munro, 1995a, 1995b, 1997).

However, according to Miller (1956) one thing to be attentive is that there might be a limit of apprehension capacity in the human brain setting up restraints of discerning no more than seven elements at once. More detailed rationales are as follows:

> Seven-point rating scales for a long time have been used to one-dimensional judgments, on the intuitive basis that trying to rate into finer categories does not really add much to the usefulness of the ratings.
>
> (Miller, 1956, p. 3).

It is subsequently presumed that as a human mind can completely process given information in the limited memory span, or within the seven categories, it would be desirable to select 7 point Likert scale to be used in the research. In order to reduce the burden on raters under the rating process with no hindrance of memory space the present study is to 7 point Likert scale instead of 9 point Likert scale.

3.2 Data Collection and Analysis

The recording data collected from the experiment were analyzed and the duration of each syllable was measured by Praat. When dividing the consecutive speech into separate segments, there might arise a problem that the researcher makes a mistake with regard to the possibility to commit inconsistent identification which might occur during the segmentation (Peterson & Lehiste, 1960). As FAVE-align program (Rosenfelder *et al*., 2011) automatically extracts the duration of each consonant and vowel, formants, pauses, etc., the present study applied this program to the data coding in order to reduce the risk of making errors and objectify the process of segmentation compared to analysis manually conducted. However, several cautions when adopting the segmentation process accompanied with the Fave-align program are to be considered.

The basic principle to detect each phone is to map the transcribed words in a Praat script into a phone sequence (e.g. from grapheme to phoneme) by means of a SCOTUS speech corpus. A Hidden Markov Model[2] was selected as a

[2] The successive set of frames from the speech signal is analyzed and less than 10 ms is found as a unit. The frames along with phonemes are aligned and determined through the Viterbi algorithm, which detects the most likely acceptable sequence of hidden elements from the given data and the acoustic

phone detector to analyze phonetic boundaries.

There can exist some challenges to be solved in order to validate the forced-alignment techniques: it might misinterpret the given acoustic data by the possible orthographic fallacy, speech variables, and incomplete transcripts of disfluencies. As the given recording materials consist of the controlled speech data, incorrect orthography can not only hardly appear while processing the acoustic analysis, but also incorrect transcripts can be attuned by the manual scrutiny. It can be implied that the chance to encounter the flaws specified above can be decreased in read speech, hence, leading the present research to employ this acoustic alignment program so as to conduct the segmentation process on the purpose of achieving both technical and manual segmentation when dividing each sentence.

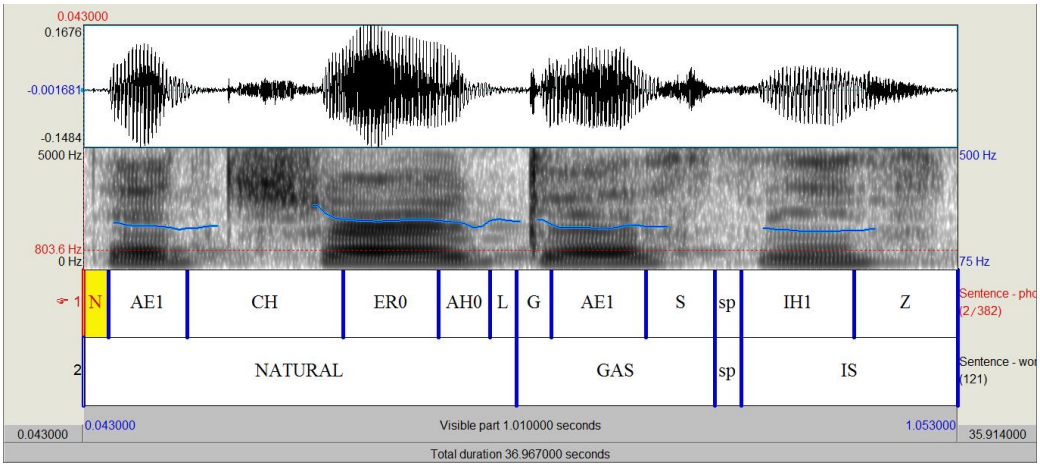
Appendix C shows the Praat script to conduct the alignments. When the whole information was demarcated, vowel duration was selectively extracted by vowel duration logger script in Appendix D. The tier 1 displays the ARPAbet of input transcriptions as shown in Figure 4. The benefit of The ARPAbet is that it also includes stress digits for all vowels, which facilitates recognizing the primary stress in utterances. The ARPAbet system can be found in Appendix E.

---

model represented by the HMMs.

After all the process was conducted the researcher double checked the aligned data to make sure if automatic alignments and vowel duration extraction were adequately performed. The final output of segmentation is presented in Figure 3.1.

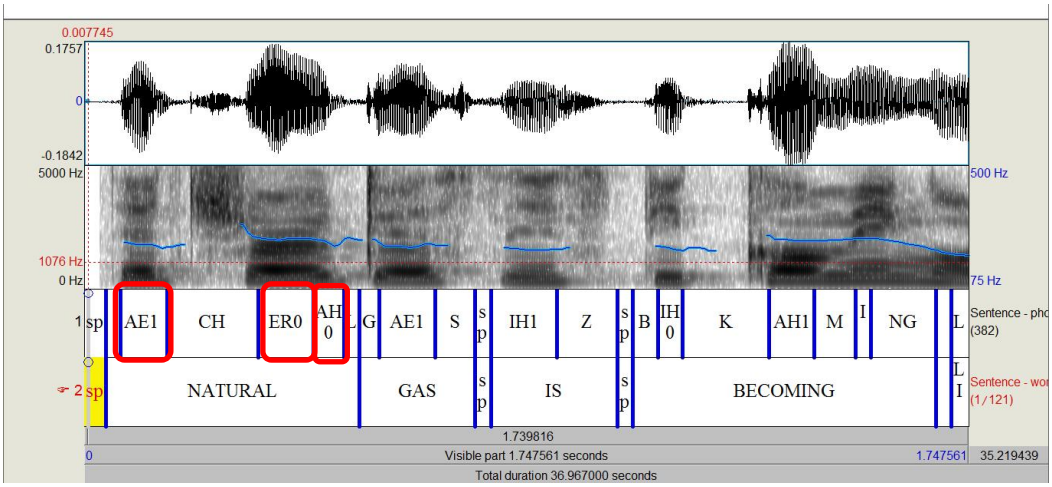
Figure 3.1
Segmentation of Speech Data by Fave-align



*Tier 1 shows the segment alignment of speech signal by the FAVE-align program.

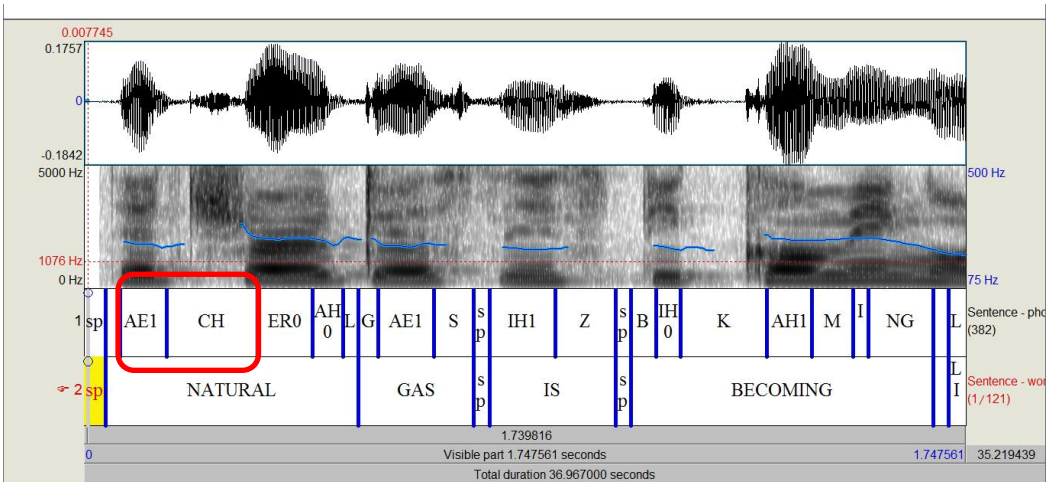
With regard to the different qualification of the syllable length, the present study examined if there are any differences in English speech rhythm between native speakers and Korean learners of English. Owing to the dispersed vocalic intervals shown in Figure 3.2, the vowel duration logger script extracted all the vowel duration set. The retained vowel duration data were substituted to the nPVI-V formula.

Figure 3.2
Vocalic Intervals arranged by Fave-align Program



If syllable-timed languages obey the rule of the speech rhythm, a lower nPVI-V and a higher nFPVI would be expected to occur. As stress, in stress-timed languages, assumes a key role to make an isochronous unit (i.e., foot) the present study applied a foot unit to the nPVI. The present research adapted the traditional definition of the foot, so that the foot duration could be measured from the onset of the stressed-syllable to the offset of the unstressed-syllable which is placed right before the next stressed-syllable. Figure 3.3 shows the arrangement regarded as the foot unit.

Figure 3.3
Foot arrangement based on the traditional definition



SPSS 21 (IBM SPSS Statistics for Windows, 2012) was adopted for the statistical analysis of the data. Two different data set: a) vocalic intervals for nPVI-V and b) the foot unit for nFPVI were compared respectively using the independent sample t-test. Considering the proficiency level of the Korean speakers, the results of nPVI-V and nFPVI were also compared by the independent sample t-test. As for the relation between speech rhythm and comprehensibility, the linear regression test was implemented.

CHAPTER 4
RESULTS AND DISCUSSION

This chapter provides the results of the study based on the data analysis. Each section is explained based on the three main research questions proposed in Chapter 1. 4.1 explains how speech rhythm based on vocalic intervals is different between the NS and the NNS. 4.2 explores whether an isochronous unit in English, foot, can show the difference between the two speaking groups. Section 4.3 demonstrates the proficiency effects on the two rhythm metrics in the speech of Korean learners of English. Section 4.4 describes the effects of sentence stress on comprehensibility of speech rhythm.

4.1 nPVI-V between NNS and NS

The goal of the present research was to demonstrate rhythmic differences between Korean learners of English and native speakers of English. Considering the widely accepted dichotomy derived from Pike (1945), it is natural to categorize English as a stress-timed language and Korean as a syllable-time language. As to the noticeable distinction, there is a tendency to assign intervals of stress evenly in stress-timed languages resulting in an inconsistency

of syllable length, while each syllable tends to retain the duration uniformly in syllable-timed languages (Abercrombie, 1967).

In Figure 4.1, the average value of nPVI-V between the non-native speakers and the native speakers were presented. As syllable duration could vary in stress-timed languages, the NS group showed the much higher result than the NNS group.

Figure 4.1
Mean Value of nPVI-V between NNS and NS

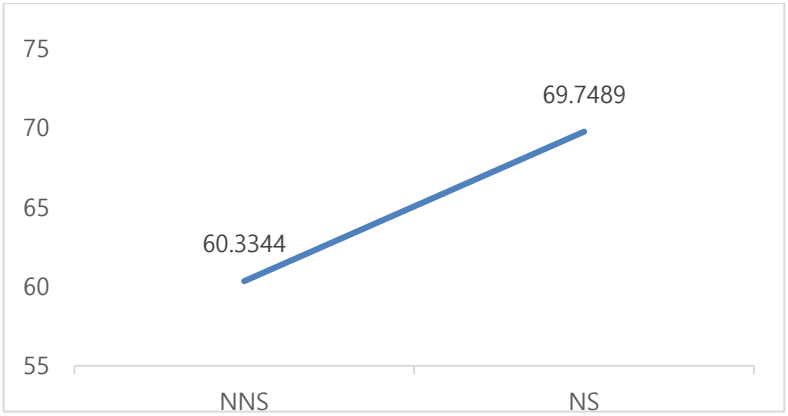


Table 4.1 shows the summary of the independent samples t-test which was conducted to compare two mean values of both groups. As there was a higher mean value in NS (M=69.7489, SD=11.459) than in NNS (M=60.3344, SD=15.171), there was significant difference between NK and NE groups [t(214)=-3.097, p=.002, <.01], which provides an indication that the distinction

of the speech rhythm between the two groups is statistically significant.

Table 4.1
Independent samples T-Test Summary of nPVI-V

	Mean		SD		95% CI for Mean Difference		t	Sig.
	NK (n=189)	NE (n=27)	NK	NE	Lower Bound	Upper Bound		
nPVI-V	60.3344	69.7489	15.171	11.459	-15.40423	-3.42476	-3.098	.002*

Discrimination between the NK and the NE groups was confirmed: the NK group did not utilize the stressed and unstressed syllable alternation, whereas the NE group demonstrated the obeisance on the speech rhythm. This result conformed to the findings of the previous researches (Grabe & Low, 2000; Grabe & Low, 2002; White, *et al*., 2007; Gibbon & Gut, 2001; Deterding, 2001).

In the major findings of the first experiment, it seems that the acoustic figure of the speech rhythm based on the vocalic intervals between the Korean speakers of English and the native speakers is significantly different. It can be assumed that the present research is in concord with the previous research (Grabe & Low, 2002; Asu & Nolan, 2006; Mok & Lee, 2008). The tendency of the low nPVI-V for Korean speakers and of the high nPVI-V for the native speakers might be supported in a way that Korean belongs to syllable-timed than stress-timed languages.

4.2 nFPVI between NNS and NS

As the foot unit was applied to the PVI formula, there was an expectation that there would be meaningful difference of nFPVI as much as nPVI-V between NNS and NS. Although Figure 4.2 shows a slight difference between the two groups, mean value between NNS (M=49.279, SD=14.086) and NS (M=49.57, SD=14.35) was meager. In the same line with the findings with Asu and Nolan (2006), there was no statistically significant differences between NNS and NS in nFPVI measurement [t(33.6)=-.100, p=.920] as shown in Table 4.2.

Figure 4.2
Mean Value of nFPVI between NNS and NS

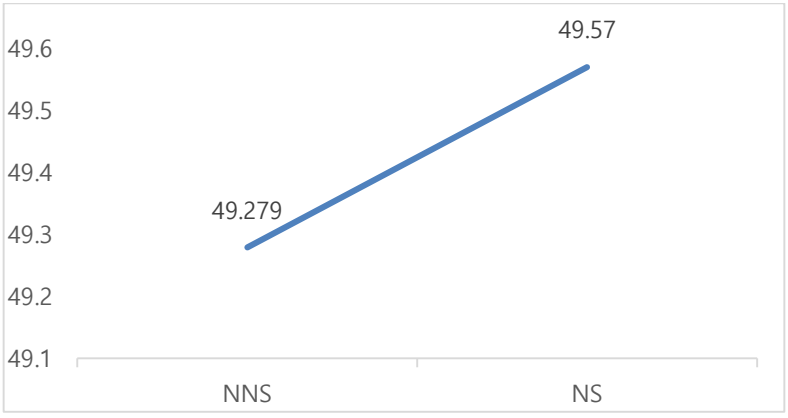


Table 4.2
Independent samples T-Test Summary of nFPVI

	Mean		SD		95% CI for Mean Difference		t	Sig.
	NNS (n=189)	NS (n=27)	NNS	NS	Lower Bound	Upper Bound		
nFPVI	49.279	49.57	14.086	14.35	-6.01618	5.43438	-.100	.920

The weak distinction of the average nFPVI scores appeared to be attributed to the same tendency of approaching the foot isochrony between the two groups. Asu and Nolan (2006) concluded that almost the same scores of the nFPVI but the huge discrepancy of nPVI-V scores between the groups might be on account of other factors taken to make the speech rhythm: the native speakers of English were likely to use the unstressed syllable reduction, while the Estonian learners of English were presumed to reduce the syllables regardless of English stress patterns. As Estonian categorically belongs to syllable-timing like Korean, it can apply the same explanation to the present research that Korean participants utilized a shortening in any syllables which could make the comparably regular feet in spite of the significant difference of vocalic intervals.

4.3 Proficiency Effects on Speech Rhythm

4.3.1Proficiency Effects on nPVI-V

Few research has held onto the claim that the speech rhythm can be various in terms of the speaker's proficiency level. Nonetheless, Stockmal *et al*., (2005) proved that advanced and low level speakers had different results according to the rhythm varieties. The research provided that low level speakers showed a higher variability than advanced speakers, indicating the beginners might be confused to articulate the longer vowels in a position where the sentence stress must be located.

On the purpose of unveiling effects of the proficiency level on the speech rhythm, the present study categorized the two groups: more proficient and less proficient groups. As the numerical figure of the statistical analysis in the present research demonstrated both groups were significantly different, there might be a huge gap between the groups especially in an aspect that the more proficient group would have a higher value in the nPVI-V than the less proficient group which was the ideal conclusion of Stockmal *et al* (2005).

Focusing on nPVI-V, the prospect conceivable here was that there might be the higher value in the MP and the lower value in the LP group. To compare

the results of the two groups, the independent samples T-Test was conducted. Figure 4.3 shows mean values of the MP and the LP groups. As the MP group attained the lower mean value (M=56.9257) than the LP group (M=61.6979), it seemed that the LP group could utter the English sentences more rhythmically than the MP group.

Figure 4.3
Mean Value of nPVI-V in MP, LP and NS

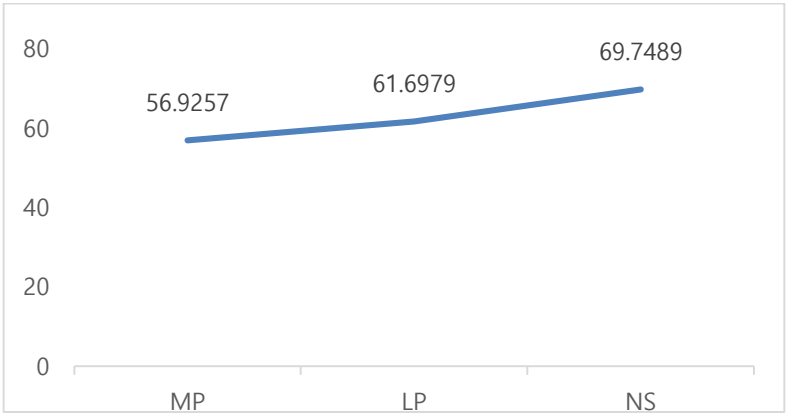


Table 4.3
Independent samples T-Test Summary of nPVI-V based on Proficiency Level

	Mean		SD		95% CI for Mean Difference		t	Sig.
	MP (n=54)	LP (n=135)	MP	LP	Lower Bound	Upper Bound		
nPVI-V	56.9257	61.6979	14.74100	15.17947	.02287	9.52136	1.993	.049

The Table 4.3 shows the summary of the analysis. The results of this test indicated that the difference between the two proficiency groups is statistically significant [t(100.4)=1.993, p=.049, <.05]. The results found in the present study, in which the difference between the proficiency levels existed, corresponded to the findings in Stockmal *et al*. (2005). It might be juxtaposed to the possibility that the less proficient Korean speakers of English misled themselves to elongate the vowel duration when the sentence stresses were needed in order to satisfy their English speech rhythm.

4.3.2 Proficiency Effects on nFPIV

Whether the different proficiency level influences the result of the nFPVI measure was another question to be examined. As for nFPVI, the MP group was expected to have a lower value than the LP group in accordance with the supposition that the proficient speakers of English might be able to maintain the isochronous feet better than the less proficient speakers.

Shown in Figure 4.4, the MP group attained the lower mean value (M=47.1259) than the LP group (M=50.1399), which is lower than the NS mean value (M=49.57). It can be elicited that the MP group complied with the timing

regularity to meet the isochrony in English.

Figure 4.4
Mean Value of nFPVI in MP, LP, and NS

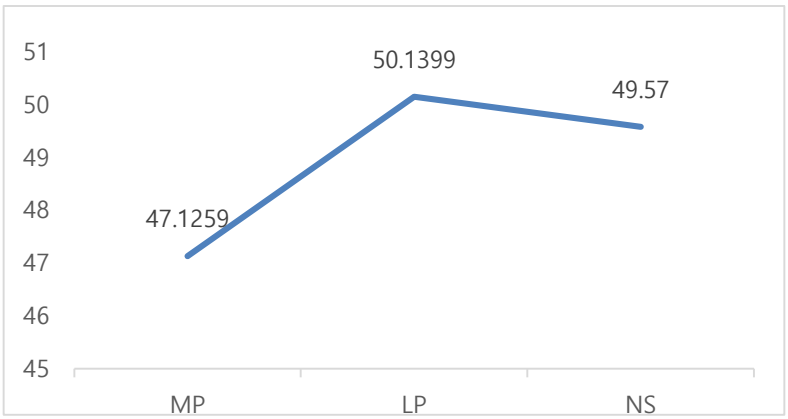


Table 4.4 shows the summary of the independent t-test which was implemented to compare the nFPVI between the MP and the LP group. Even though the mean value was different in line with the anticipation that the MP would have a lower nFPVI than the LP group, it failed to manifest the statistical significance [t(187)=1.332, p=.185].

Table 4.4
Independent samples T-Test Summary of nFPVI based on Proficiency Level

	Mean		SD		95% CI for Mean Difference		t	Sig.
	MP (n=54)	LP (n=135)	MP	LP	Lower Bound	Upper Bound		
nFPVI	47.1259	50.1399	12.27155	14.70297	-1.45102	7.47887	1.332	.185

The nFPVI scores between both speaking groups showed almost the same value as there was no statistically significant difference. It might be impetuous to draw a conclusion that the utterances of languages in connected speech do not have rhythm when it comes to the foot unit. Rather, it might be plausible that nFPVI cannot detect the rhythm distinction properly. For example, Asu and Nolan (2009) found that the result manifested there was a higher nFPVI in Mexican Spanish than in Castilian. As Mexican Spanish was supposed to be more stress-timed than Castilian, for the former was assumed to be connected with American English varieties whereas the latter with European varieties, the findings remained to be more interpreted in future research.

In total, the tendency that there was a huge gap in the nPVI-V along with the proficiency level and the different language backgrounds, and almost the equal scores of the nFPVI regardless of either the proficiency level or the different speech groups, might need to be in question. According to Asu and Nolan (2006), in fact, it was speculated that there would be the ideal results of

the nFPVI: the lower nFPVI and the higher nPVI-V for stressed-timing and the higher nFPVI and the lower nPVI-V for syllable-timing. If the MP group is postulated to have a command of English speech rhythm like a native speaker better than the LP group, enabling them to employ a slight regular foot than the LP group, it might be advisable to suggest that the nFPVI metric succeeded in differentiating the scores between the groups, which would be appropriate to determine the rhythm structure by the foot unit. Much caution, still, needs to be put forward as the MP group scores of nFPVI were not critical to postulate the presupposition presented above. The nPVI-V scores were also not impressively higher than those of the LP group in spite of the lack of rationale to explain why the less proficient group acquired the higher scores of nPVI-V.

One clumsy but probable clue to grasp this phenomenon is that the different strategy to articulate the English rhythm, in practice, might be applied to the speakers' speech data (e.g. lengthening the stressed syllables or lenition of the consonants cluster), contributing to the ambiguous explication of nPVI. In accordance with the conclusion of Asu and Nolan (2006, 2009), nFPVI not only needs to be improved more as a new tool to unveil the identity of the speech rhythm but also requires more empirical backgrounds to bolster its rationality.

4.4 Comprehensibility based on Sentence Stress and Speech Rhythm

What decisively marked off the present study from the previous research of rhythm was that there was an attempt to conduct both objective and subjective measurements of speech rhythm altogether. Although the computed segmentation of speech information and a numerical value corresponding to nPVI contributed to objectifying the measurement speech rhythm, there were more requests to compensate the limits of the rhythm measures. Many researchers suggested that numerous types of rhythm metrics might have deficiency to uphold the speech rhythm hypothesis in that the results of individual experiments conducted by different researchers have been inconsistent (Mok & Lee, 2008; Couper-Kuhlen, 1993; Roach, 1982).

As many rhythm metrics did not satisfy the exact classification, there have appeared the view that rhythm is a matter of perception rather than acoustically measureable entity. Tan and Low (2014) suggested that perceptual experiments might need to be entailed so that the acoustic findings could be enriched on the basis of speech rhythm. Nazzi and Ramus (2003) found the listeners were able to discriminate language rhythm when performed a perceptual test.

Bound up with the perceptual recognition, but more with the extent to

difficulty processing the given acoustic data, comprehensibility will be the major focus to be applied to examine whether the speech rhythm affects the overall comprehensibility. In order to confirm the relation between the sentence stress and comprehensibility the native speakers were asked to rate the learners' speech data with a 7 Likert-scale based on two standards: a) whether the sentence stresses were well obeyed, and b) whether the sentence stresses produced by the Korean learners helped the raters easily to understand the sentences as well as follow the speech rhythm.

Table 4.5
Linear Regression Analysis Summary: The Relationship between Sentence Stress and Comprehensibility

	Independent Variable	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
Comprehensibility	(Constant)	2.528	.130		19.429	.000
	Sentence Stress	.591	.031	.609	19.130	.000

R^2 = .370, Adjusted R^2= .369, F= 365.946, p= .000,

R^2 Std. Error of the Estimate = .87446

As shown in Table 4.5, a simple linear regression was calculated to predict comprehensibility based on sentence stress. The sentence stress was set as an independent variable and comprehensibility as a dependent variable. It was statistically approved that the sentence stress had effects on comprehensibility [β=.609, t(622)=19.130, p<.001]. The results of the regression were the statistically significant [R^2=.370, F(1,622)= 365.946, p<.000]. It suggests that 37% of the variance in comprehensibility can be explained by the sentence stress. It is, thus, to be expected that the sentence stress the Korean learners uttered has effects on the comprehensibility when the native speakers judge the difficulty in understanding the speech rhythm.

CHAPTER 5
CONCLUSION

This chapter draws a conclusion with respect to the results and discussions in the previous chapter. Section 5.1 summarizes major findings of the present research and offer pedagogical implications. Section 5.2 indicates the limitations of the current study. Section 5.3 presents implications and suggestions for future research.

5.1 Summary of the Study and Pedagogical Implications

The present study mainly aimed to explore the speech rhythm between Korean high school learners of English and native English speakers by using rhythm measures. The second research questions was to examine the proficiency effect of Korean learners on English speech rhythm. In order to answer these two inquiries, nPVI-V and nFPVI were selected to measure the speech rhythm based on the vocalic intervals and the foot unit. The last objective was to clarify if the sentence stress really affects the rhythm comprehensibility and can be regarded as one of the attributes in the speech rhythm perception. The major findings of the first experiment are presented in the following manner.

First, it was found that the acoustic figure of the speech rhythm based on the vocalic intervals between the Korean speakers of English and the native speakers was significantly different, which is in concord with the previous research (Grabe & Low, 2002; Asu & Nolan, 2006; Mok & Lee, 2008). The tendency of the low nPVI-V for Korean speakers and of the high nPVI-V for the native speakers can be supported in a way that Korean belongs to syllable-timed than stress-timed languages. In contrast, the nFPVI scores between the two speaking groups failed to show the statistically significant difference, which is in line with the result of Asu and Nolan (2006).

Second, the present study also analyzed the English speech rhythm comparing the two different proficiency levels of the Korean learners. The result showed that the less proficient group had the higher scores than the more proficient group. Even though the gap in nPVI-V between the two groups was statistically significant, it did not support the presumption that the more proficient group would reach the more native-like scores than the less proficient group.

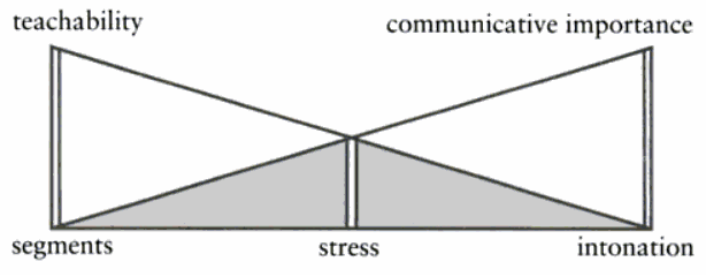
Lastly, the comprehensibility rating was conducted to manifest the relationship between the sentence stress and the speech rhythm. It was confirmed that there was a correlation of the sentence stress with the speech rhythm. In other words, the better the sentence stress was placed, the better the speech

rhythm was grasped, which indicates that a good stress arrangement in sentences can make speech rhythm more comprehensible.

At the moment in existing teaching practices, there is a lack of research related to practical pronunciation pedagogy, which has yet to be sufficiently integrated into discussion of the four language skills (e.g. listening, speaking, reading, and writing). Elliott (1997) mentioned that, with the exception of pronunciation, most of these related language skills have been explored in communicative language settings. As a result, pedagogical practices related to pronunciation instruction are still underdeveloped and require further attention and inquiry.

Dalton & Seidlhofer (1994) underlines, among segmental and suprasegmental features, that stress can be one of the effective factors in pronunciation teaching with regard to teachability and learnability in a class boundary. Overall meaningful messages are primarily delivered by suprasegmentals in communication, so that it might be pinpointed that prosody can provide the listener with the speaker’s point of view. Despite the intimate connection between prosodic features and communication, there remains a problem that suprasegmentals are too difficult to teach compared to segmental elements.

Figure 5.1
Teachability and Learnability



(Dalton & Seidlhofer, 1994, p. 73)

As shown in Figure 5.1, Dalton and Seidlhofer (1994) illustrated a continuum of segmentals and suprasegmentals based on teachability and communicative importance (learnability). As stress can play a bridge-building role of the two ends, not only does it seem to be effective for both teachers and learners to approach the pronunciation learning in class, but also it might help students to achieve both acoustic and phonological features altogether.

5.2 Limitations and Suggestions for Future Research

The current study attempted to shed light on the numerical and the perceptual entity of the speech rhythm comparing the Korean EFL high school learners of English and the native English speakers. It also examined the rhythm

performance according to the different proficiency levels of the Korean students. Even though the current research provided the several noticeable findings corresponding to the previous research of the speech rhythm, there remain several limitations.

First, there still exist controversial issues about the trend to discriminate the rhythm class of languages. Lehiste (1977) reported some studies failed to control a variety of factors such as syntactic boundaries or syllable-final lengthening, etc. As most of the rhythm measures did not yield a transparent conclusion, some researchers began to regard Isochrony as a perceptually recognized rhythmic flow, rather than an acoustic realization that can be numerated (Roach, 1982; and Dauer, 1983). It can be assumed, in other words, that isochrony is the authentic phenomenon of the perceived speech bearing an acoustically illusory presence (Couper-Kuhlen, 1993). For this reason, the convention of dichotomized speech rhythm (i.e. stress- and syllable-timing) is making a concession to a concept of a continuum (Couper-Kuhlen, 1993; Crystal, 1996; Dauer, 1983; Ramus, 2002; Miller, 1984).

Second, the speech recording implemented in the present study hardly reflected connected speech as the participants read the reading sheet instead of making impromptu utterances. As the reading sentences were sophisticatedly manipulated in terms of the number of syllables and words, there would be

deficiency to examine the authoritative characterization of connected speech. It might be desirable that the future studies collect data from a natural but spontaneous speech, so that it can help to epitomize how speakers produce their speech rhythm.

Third, the present research concentrated on finding out the acoustic difference between and native English speakers and the Korean learners of English by means of nPVI-V and nFPVI, which was originally derived from PVI by Grabe and Low (2002). As various new types of PVI were developed to investigate speech rhythm in the different aspects, there have been miscellaneous attempts to apply PVI to discriminate languages according to the rhythm class. Ferragne (2013) also adjusted the original formula of PVI by exploiting mean values of syllable intensity and pitch for a new design, which succeeded in distinguishing accents better duration. Despite the versatile availability of the PVI, its broader utilization in the empirical research is meager in practice. Therefore, it might be advisable to apply a range of newly devised PVI measures to the field of speech rhythm for future research.

REFERENCES

김진완 외. (2015). *실용영어 1.* 서울: 천재교육.

교육과학기술부. (2011). *영어과 교육과정.*

한종임. (2015). *영어음성학과 발음지도(개정판).* 한국문화사.

이성엽. (2013). *영어발음&리듬훈련노트.* 서울: 길벗이지톡.

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press.

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(3), 714.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension*. *Language learning, 38*(4), 561-613.

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

Ashby, M., & Maidment, J. (2005). *Introducing phonetic science*: Cambridge University Press.

Asu, E. L., & Nolan, F. (2005). *Estonian rhythm and the pairwise variability index.* Paper presented at the Proceedings of Fonetik 2005.

Asu, E. L., & Nolan, F. (2006). *Estonian and English rhythm: a two-dimensional quantification based on syllables and feet.* Paper presented at the Proceedings of Speech Prosody.

Breitkreutz, J. A., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation Teaching Practices in Canada. *TESL Canada Journal, 19*(1), 51-61.

Caspers, J. (2010). The Influence of Erroneous Stress Position and Segmental

Errors on Intelligibility, Comprehensibility and Foreign Accent in Dutch as a Second Language. *Linguistics in the Netherlands*(27), 17-29.

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation*. New York: Cambridge University Press.

Couper-Kuhlen, E. (1993). *English speech rhythm : form and function in everyday verbal interaction*. Amsterdam ; Philadelphia: J. Benjamins.

Crystal, D. (1996). The past, present and future of English rhythm. *Speak Out, 18*, 8-13.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech, 40*(2), 141-201.

Dalton, C. & Seidlhofer, B. (1994). *Language teaching: pronunciation*. Oxford University Press.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient forΔ C. *Language and language-processing*, 231-241.

Dellwo, V., Diez, F. G., & Gavalda, N. (2009). *The development of measurable speech rhythm in Spanish speakers of English.* Paper presented at the Actas de XI Simposio Internacional de Comunicacion Social.

Dellwo, V., & Wagner, P. (2003). Relationships between rhythm and speech rate.

Derwing, T. M., & Munro, M. J. (1997). Accent, Intelligibility, and Comprehensibility: Evidence from Four L1s. *Studies in second language acquisition, 19*(1), 1-16.

Derwing, T. M., & Rossiter, M. J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning, 13*(1), 1-17.

Elliott, A. R. (1997). On the Teaching and Acquisition of Pronunciation within a Communicative Approach. *Hispania, 80*(1), 95-123.

Ferragne, E., Ballier, N., Díaz-Negrillo, A., & Thompson, P. (2013). Automatic suprasegmental parameter extraction in learner corpora. *Automatic treatment and analysis of learner corpus data*, 151-168.

Field, J. (2005). Intelligibility and the Listener: The Role of Lexical Stress. *TESOL Quarterly, 39*(3), 399-424.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244-249.

Gibbon, D., & Gut, U. (2001). *Measuring speech rhythm.* Paper presented at the INTERSPEECH.

Giegerich, H. J. (1992). *English phonology: An introduction*: Cambridge University Press.

Gilakjani, A. P. (2012). The Significance of Pronunciation in English Language Teaching. *English Language Teaching, 5*(4), p96.

Gow Jr, D. W., & Gordon, P. C. (1993). Coming to terms with stress: Effects of stress location in sentence processing. *Journal of psycholinguistic research, 22*(6), 545-578.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology, 7*(515-546).

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 201-223.

Jang, T.-Y. (2009). *Automatic assessment of non-native prosody using rhythm metrics: Focusing on korean speakers' english pronunciation.* Paper presented at the Proc. of the 2nd International Conference on East Asian Linguistics.

Jenkins, J. (2004). 5. RESEARCH IN TEACHING PRONUNCIATION AND

INTONATION. *Annual review of applied linguistics, 24*, 109-125.

Ladefoged, P. (2006). *A course in phonetics* (Vol. 5th ed). Australia ; Boston, MA: Thomson, Wadsworth.

Lee, H. Y. (1997). *Korean prosody*. Seoul: Korea Research Center.

Lehiste, 1. (1977). Isochrony reconsidered. *Journal of Phonetics 5*, 253-263.

Lloyd James, A. (1940). *Speech signal in telephony*. London: Pitman & Sons.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review, 63*(2), 81.

Miller, M. (1984). On the perception of rhythm. *Journal of Phonetics*, 12, 75–83.

Mok, P., & Lee, S. I. (2008). *Korean speech rhythm using rhythmic measures.* Paper presented at the Proc. of the 18th Intern. Congress of Linguists (CIL18), Seoul, Korea.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning, 45*(1), 73-97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech, 38*(3), 289-306.

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

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Commun., 41*(1), 233-243. doi: 10.1016/s0167-6393(02)00106-1.

Nolan, F., & Asu, E. L. (2009). The pairwise variability index and coexisting

rhythms in language. *Phonetica, 66*(1-2), 64-77.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustical Society of America, 119*(5), 3034-3047.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693-703.

Pike, K. L. (1945). The intonation of American English.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265-292.

Roach, P. (1982). On the distinction between 'stress-timed'and 'syllable-timed'languages. *Linguistic controversies*, 73-79.

Rosenfelder, Ingrid; Fruehwald, Joe; Evanini, Keelan and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite. http://fave.ling.upenn.edu.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842-854.

Steele, J. (1779). *Prosodia Rationalis: Or, An Essay Towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols*: J. Nichols, sold by T. Payne. [Reprinted, Meston: The Scolar Press, 1969]

Stockmal, V., Markus, D., & Bond, D. (2005). Measures of native and non-native rhythm in a quantity language. *Language and speech, 48*(1), 55-63.

Tan, R. S. K., & Low, E.-L. (2014). Rhythmic Patterning in Malaysian and Singapore English. Language and speech, 57(2), 196-214.

Tyler, L. K., & Warren, P. (1987). Local and global structure in spoken language

comprehension. *Journal of Memory and Language, 26*(6), 638-657.

Wagner, A. Description of Polish speech rhythm using rhythm metrics and the time-delay approach: A comparative study.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of phonetics, 35*(4), 501-522.

Winters, S., & O'Brien, M. G. (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication, 55*(3), 486-507.

Wong, R., Languages, E. C. o., & Linguistics. (1987). *Teaching pronunciation*. Englewood Cliffs, N.J.: Prentice-Hall.

APPENDICES

APPENDIX A.

Nine Manipulated Sentences

Sentence Length	# of Words	# of Syllable	Sentence
Short (Less than 10 syllables)	6	12	Watermelon isn't in a fruit category.
	8	11	I prefer to wait for the perfect person.
	6	9	Kevin wanted to solve the problem.
Medium (Less than 15 syllables)	8	16	The flight to America prepared for emergency landing.
	9	16	Local department stores are full of excited shoppers today.
	9	14	My parents decided to give me some advice today.
Long (Less than 20 syllables)	12	22	Natural gas is becoming a limited energy source to make a battery.
	14	20	My father and grandpa used to smoke a pack of cigarettes in the afternoon.
	12	24	I have studied computer animation to become an animator in Disney studio.

APPENDIX B.

Three Sentence Fillers

Fillers
Athena disguised herself as an old woman and appeared at Arachne’s house.
Medusa was a beautiful woman and very proud of her beauty.
People with too much pride are always punished.

APPENDIX C.

Vowel Alignment Script by FAVE-align

```
########################################################################
## This Praat script exports orthographic transcriptions in Praat      ##
## to a format suitable as input to the FAVE-align forced aligner      ##
## (http://fave.ling.upenn.edu/FAAValign.html).                        ##
## The transcription will be converted to a 5-column tab-delimited .txt file ##
## as outlined in the instructions on the FAVE web site.               ##
##                                                                     ##
## To run this program, select the TextGrid containing the transcriptions, ##
## open this script, and select "Run" > "Run".                         ##
##                                                                     ##
## This script was written by Ingrid Rosenfelder,                      ##
## last modified October 31, 2011                                      ##
########################################################################

## get TextGrid name
info$ = Info
filename$ = extractLine$(info$, "Object name: ")
outfile$ = filename$ + ".txt"
## ask the user before overwriting exiting file
if fileReadable(outfile$)
        pause 'filename$'.txt already exists.    Overwrite?
        deleteFile(outfile$)
endif

## extract transcription info and write to file
n_tiers = Get number of tiers
for tier from 1 to n_tiers
        tiername$ = Get tier name... 'tier'
        n_intervals = Get number of intervals... 'tier'
```

```
        for interval from 1 to 'n_intervals'
                start = Get start point...    'tier' 'interval'
                end = Get end point... 'tier' 'interval'
                label$ =    Get label of interval... 'tier' 'interval'
                if label$ <> ""
                        fileappend 'outfile$'
'tiername$"tab$"tiername$"tab$"start"tab$"end"tab$"label$"newline$'
                endif
        endfor
endfor

echo Written transcription in FAVE-align input format to file 'outfile$'.
```

APPENDIX D.

Vowel Duration Logger Script

```
##duration logger.praat
##written by Katherine Crosswhite
##modified by Mark Antoniou

##What does this script do?
##Outputs the duration of all intervals marked in tier 1 with non-null lables.
Durations, in milliseconds, will be written to a file called "Durations.txt", which
you will be able to find in the same directory holding your sound files after you
run the script.

##    Specify the directory containing your sound files in the next line:
form Enter directory containing TextGrids:
#Be sure not to forget the slash (Windows: backslash, OSX: forward slash) at the
end of the directory name.
        sentence Directory G:\
endform

#Now we will do some prep work to get your log file ready.    The first thing I
usually do is make sure that I delete any pre-existing variant of the log:
filedelete 'directory$'durations.txt

##    Now I'm going to make a variable called "header_row$", then write that
variable to the log file:
header_row$ = "Filename" + tab$ + "phoneme" + tab$ + "Duration (ms)" +
newline$
header_row$ > 'directory$'duration_log.txt

##    Now we make a list of all the text grids in the directory we're using, and put
```

```
the number of filenames into the variable "number_of_files":
Create Strings as file list...    list 'directory$'*.TextGrid
number_files = Get number of strings

# Then we set up a "for" loop that will iterate once for every file in the list:
for j from 1 to number_files

        # Query the file-list to get the first filename from it, then read that file in:
        select Strings list
        current_token$ = Get string... 'j'
        Read from file... 'directory$"current_token$'

        # Here we make a variable called "object_name$" that will be equal to the
filename minus the ".wav" extension
        object_name$ = selected$ ("TextGrid")

        # Now we figure out how many intervals there are in tier 1, and step
through them one at a time.
        # If an intervals label is non-null, we get its duration and write it to the log
file.
        number_of_intervals = Get number of intervals... 1
        for b from 1 to number_of_intervals
                interval_label$ = Get label of interval... 1 'b'
                if interval_label$ <> ""
                        begin_vowel = Get starting point... 1 'b'
                        end_vowel = Get end point... 1 'b'
                        duration = (end_vowel - begin_vowel) * 1000
                       fileappend "'directory$'duration_log.txt"
'object_name$"tab$"interval_label$"tab$"duration:3"newline$'
                endif
        endfor

        #    By this point, we have gone through all the intervals for the current
        #    textgrid, and written all the appropriate values to our log file.    We are
now ready to go on to
```

```
        #    the next file, so we close can get rid of any objects we no longer need,
and end our for loop
        select all
        minus Strings list
        Remove
endfor

# And at the end, a little bit of clean up and a message to let you know that it's all
done.
select all
Remove
clearinfo
printline All files have been processed.
printline The durations have been output to 'directory$'
```

APPENDIX E.

ARPAbet Chart

Vowels			Consonants		
Phoneme	**Example**	**Transcription**	**Phoneme**	**Example**	**Transcription**
AA	bot	B AA T	B	be	B IY
AE	bat	B AE T	CH	cheese	CH IY Z
AH	but	B AH T	D	day	D EY
AO	bought	B AO T	DH	that	TH AE T
AW	bout	B AW T	F	fee	F IY
AY	bite	B AY T	G	go	G OW
EH	bet	B EH T	HH	he	HH IY
ER	bird	B ER D	JH	just	JH AH S T
EY	bait	B EY T	K	key	K IY
IH	bit	B IH T	L	late	L EY T
IY	beat	B IY T	M	me	M IY
OW	boat	B OW T	N	knee	N IY
OY	boy	B OY	NG	sing	S IH NG
UH	put	P UH T	P	pay	P EY
UW	boot	B UW T	R	read	R IY D
			S	sea	S IY
			SH	she	SH IY
			T	tea	T IY
			TH	thanks	TH AE NG K S
			V	vain	V EY N
			W	we	W IY
			Y	yes	Y EH S
			Z	zoo	Z UW
			ZH	pleasure	P L EH ZH ER

Stress is indicated by digits following the stressed vowels. There are three levels of stress:

Value	Level of stress
0	no stress
1	primary stress
2	secondary stress

Example:

in	AH0 N, IH1 N
the	DH AH0, DH AH1, DH IY0
dictionary	D IH1 K SH AH0 N EH2 R IY0
stress	S T R EH1 S
is	IH1 Z, AH0 Z
indicated	IH1 N D AH0 K EY2 T AH0 D
by	B AY1
digits	D IH1 JH AH0 T S
following	F AA1 L OW0 IH0 NG
stressed	S T R EH1 S T
vowels	V AW1 AH0 L Z

APPENDIX F.

Survey Sheet for Korean Learners

Survey Sheet

- 참가자 번호: _____
- 학년: _____학년
- 성별 □ 남자 □ 여자
- 영어 6월 모의고사 듣기/총점: _____(듣기)/_____(총점)

1. 영어사용국가 거주 및 연수 경험이 있습니까? □ 네 (2번) □ 아니오 (3번)
2. 영어사용국가 거주 및 연수 기간은 얼마입니까? _____(개월, 년)
3. 지금까지 영어공부를 한 기간은 총 얼마입니까? _____ (시간, 개월, 년)
4. 일주일에 영어로 말 하는 시간은 총 얼마입니까? _____ (분, 시간)
5. 학교에서 일주일 동안 영어 공부한 시간은 얼마입니까? _____ 시간
6. 학교에서 배우는 영어는 주로 무엇입니까? _____
7. 학교에서 영어발음을 공부한 적이 있습니까? □ 네 (8번) □ 아니오 (10번)
8. 학교에서 영어발음을 공부한 기간은 얼마입니까? _____ (시간, 개월, 년)
9. 학교에서 영어발음을 공부할 때 주로 배운 것은 무엇입니까?
   □ 자음&모음 □ 강세 □ 억양 □ 기타 ( )
10. 학교 외에서 영어공부를 한 적이 있습니까? □ 네 (11번) □ 아니오 (17번)
11. 학교 외에서 영어공부를 한 경우 어떤 곳에서 주로 도움을 받았습니까?
    □ 보습(입시)학원 □ 어학원 □ 개인과외 □ 원어민과외 □ 기타 ( )
12. 학교 외의 장소에서 영어를 공부한 기간은 얼마입니까?
    (장소가 여러 곳인 경우 해당되는 각각의 장소에 대한 기간을 적으시오.)
    장소: _____, _____ (시간, 개월, 년)
    장소: _____, _____ (시간, 개월, 년)
    장소: _____, _____ (시간, 개월, 년)
13. 학교 외의 장소에서 영어를 배울 때 주로 무엇을 배웠습니까? _____
14. 학교 외의 장소에서 영어발음을 배운 적 있습니까? □ 네 (15번) □ 아니오 (17번)
15. 학교 외의 장소에서 영어발음을 공부한 기간은 총 얼마입니까? _____ (시간, 개월, 년)
16. 학교 외의 장소에서 영어발음을 공부할 때 주로 배운 것은 무엇입니까?
    □ 자음&모음 □ 강세 □ 억양 □ 기타 ( )
17. 자신의 전반적인 영어능력을 평가한다면?
    □ 아주 나쁨 □ 나쁨 □ 보통 □ 좋음 □ 아주 좋음
18. 영어에 대한 흥미 정도는?
    □ 전혀 없음 □ 없음 □ 보통 □ 많음 □ 아주 많음
19. 자신의 영어발음을 평가한다면?
    □ 아주 나쁨 □ 나쁨 □ 보통 □ 좋음 □ 아주 좋음

감사합니다~♡

APPENDIX G.

Survey Sheet for Native Speakers

Survey

1. Participant Number: _____

2. sex: male / female

3. nationality: _____ state, USA

4. Have you ever learned other languages but for Korean? Yes ☞ 5, No ☞ 7

5. Which language?_____

6. How long have you studied other languages? _____
_____
_____

7. How long have you stayed in Korea? _____

8. How often do you speak with Koreans in English?

① never ② seldom ③ sometimes ④ often ⑤ usually ⑥ always

9. If you have learned Korean, how long do you speak in Korean?

① never ② seldom ③ sometimes ④ often ⑤ usually ⑥ always

10. How much do you think you are familiar with English accents of Korean people?

① not at all ② seldom ③ a little bit ④ on average ⑤ a lot ⑥ totally

Thank you ^^*

국 문 초 록

영어가 국제어로 대두됨에 따라 영어 발음을 학습할 때 음소체계를 정확하게 습득하는 대신 다양한 발음 양상을 고려하여 전반적인 이해도에 도움을 주는 습득에 목표를 두기 시작했다. 따라서 초분절자질이 새로운 목표를 달성하는데 필수적인 요소로 대두되었고 그 중에서도 영어강세와 리듬구현이 발음 학습에 유의미한 영향을 미칠 것으로 기대되었다. 그러나 한국인 학습자들은 영어를 발화할 때 영어리듬단위와 상충하는 리듬단위를 구사하기 쉬운데 그 이유는 한국어가 음절단위 리듬을 지니는 반면 영어는 강세단위 리듬을 지니는 것으로 구분되기 때문이다. 따라서 본 연구는 (1) 한국인 영어학습자와 원어민 사이에서 발화리듬에 차이가 있는지, (2) 좀 더 유창한 집단과 덜 유창한 한국인 집단 사이에서 발화리듬에 차이가 있는지, (3) 한국인 학습자의 문장강세가 원어민의 영어리듬 이해도에 어떤 영향을 미치는지에 대한 연구문제를 상정하여 한국인의 영어리듬이 원어민의 리듬과 얼마나 차이를 보이는지 살피고 한국인의 영어 발화리듬을 원어민이 이해하는 데 있어 영어강세가 얼마나 영향을 미치는지 확인하였다.

21명의 한국 고등학교 학습자들과 3명의 원어민이 실험에 참가

하였고 본 연구문제를 해결하기 위하여 두 개의 주된 실험이 실시되었다. 그 중에서 첫 번째 실험은 다시 두 개의 하위실험으로 나뉘었는데 첫 번째 하위실험에서는 한국인 화자와 원어민 화자의 nPVI-V 점수와 nFPVI 점수를 계산하여 발화리듬의 수치를 비교하였다. 두 번째 하위실험에서는 좀 더 유창한 집단과 덜 유창한 한국인 집단을 6명과 15명으로 각각 나누어 nPVI-V 점수와 nFPVI 점수가 어떤 차이를 보이는지 살펴보았다. 두 번째 실험에서는 원어민이 한국인 화자의 발음을 듣고 강세가 발화 리듬 이해도에 얼마나 영향을 미치는지 평가하는 실험을 실시하였다.

실험결과, 다음과 같은 유의미한 결과를 발견할 수 있었다. 첫째, 원어민과 한국인 화자 사이의 nPVI-V 점수차가 크게 벌어지는 것으로 보아 한국인 학습자들은 영어리듬을 구현하기 위해 영어의 강세를 잘 활용하지 못 하는 것으로 보였다. 둘 째, nFPVI 점수의 경우 큰 차이를 보이지 않았기 때문에 nFPVI로는 한국어와 영어의 리듬이 다른 유형으로 분류된다는 기존의 주장을 뒷받침하지 못 하는 것으로 드러났다. 영어리듬의 본질인 강세음절과 비강세음절의 규칙적인 발현을 놓고 볼 때 음보단위가 영어의 등시성을 보여준다고 여겨진다. 이러한 측면에서 원어민은 낮은 nFPVI 점수를, 한국인 화자는 높은 nFPVI 점

수를 보일 것으로 예상되지만 결과적으로 원어민과 한국인 학습자 사이에서 점수상의 큰 차이를 보이지 못 했다. 이러한 현상은 좀 더 유창한 집단과 덜 유창한 집단 사이에서도 동일하게 나타났다. 마지막으로 영어리듬에 기반을 둔 이해도 측정실험에서는 원어민이 이해도를 측정할 때 한국인 학습자의 영어강세가 발화리듬에 영향을 미치는 것으로 드러났다. 즉, 한국인 영어 학습자들이 영어를 말 할 때 영어의 리듬단위를 잘 구현하면 전체 리듬발화의 이해도가 높아지고 잘 구현하지 못 하면 전체 발화의 이해도가 낮아지는 것으로 볼 수 있다. 결과적으로 한국인 영어 학습자는 영어를 발화할 때 리듬단위가 원어민의 것과 차이가 나는 것으로 보이며 이 때 강세를 잘 구현할 수록 영어리듬의 이해도가 높아지는 것을 확인할 수 있었다. 이러한 연구결과는 한국인 학습자들의 발화리듬에 대한 이해를 제공할 뿐만 아니라 발음교육에 있어서 강세와 영어리듬단위 교육의 중요성을 시사한다.

주요어: 운율, 발화리듬, 강세기반리듬, 음절기반리듬

학번: 2011-21524