A comparative study on the production of inter-stress intervals of English speech by English native speakers and Korean speakers

Jong-Pyo Lee
Department of English, Hankuk University of Foreign Studies
zhongpiao@hufs.ac.kr

Tae-Yeoub Jang
Department of English, Hankuk University of Foreign Studies
tae@hufs.ac.kr

Abstract

This study attempts to compare the Inter-Stress Interval (ISI) patterns of English between the native speakers of English and Korean. One of the invariable results of the experiments about English speech rhythm has been that the strict concept of isochronism did not seem to exist at least in the surface phonetic level. However, the remarkable difference shown from the production experiment of the present study suggests that distinction in language rhythms, especially between English and Korean, be apparent. While the English native speakers and the proficient Korean speakers of English consistently produce rather shorter increase in ISI duration as the number of unstressed syllables located between target stressed syllables increased, the non-proficient Korean speakers of English produce a little longer one. The position of an ISI in a sentence does not seem to affect critically the duration of the ISI.

1. Introduction

Basically, the present study is based on Mochizuki-Sudo and Kiritani (1991) in its source idea. Their research has examined the production and the perception of durational pattern by Americans and by Japanese learners of English, and concluded the ISI played a role of temporal units in the production and the perception of English for Americans, but not for the non-proficient Korean speakers of English. What is meaningful about the previous study is that, in fact, it was the first effort to examine the stress-related durational pattern of ISIs by Japanese learners of English.

The stress-related durational pattern is the issue represented by isochronism in English speech rhythm; the inter-stress intervals (ISIs) tend to be equal in duration. And such concept has long been challenged by a lot of sound and speech specialists all over the world. Though the first which has named the two types of rhythms of world languages was Pike (1945), the notion that English is spoken with a stress-timed rhythm dates much more back to Steele (1775), which has devised an original idea of representing speech rhythm by means of musical notation. Meanwhile, the first one in the twentieth century who has suggested the stress-timing hypothesis in English was Jones (1960). He wrote in 1918 that in English there “is a strong tendency in connected speech to make stressed syllables follow each other as nearly as possible at equal distances.” One of the first instrumental investigations on this hypothesis was Classe (1939), which has, however, failed to testify for absolute isochrony both in production and perception. After that, although several other experiments from the point of view of production have explored such tendency of English in a variety of ways (Shen & Peterson, 1962; O’Connor, 1965, 1968; Bolinger, 1965; Uldall, 1971, 1972; Lea, 1974) they only showed a critical discrepancy in ISI durations, so that they “either rejected the claim that English is a language characterized by isochrony, or attempted to reinterpret the experimental findings to take into account the fact that perfect isochrony cannot be found in production.” (Lehiste, 1977)

In spite of all the instrumental evidence against stress-timed isochrony in English or, more broadly, the dichotomy between stress-timed rhythm and syllable/mora-timed rhythm in world languages, even Classe (1939) did not totally give up the notion of isochrony. And many other investigations remained for the underlying tendency presenting very usual phenomena in English such as compensatory shortening of a stressed vowel, which might exhibit a strong evidence for the primitive claim. Lehiste (1972), Huggins (1975), Fowler (1977), and Rakerd, Sennett & Fowler (1987) have demonstrated “foot-level shortening” in English speech rhythm. And Mochizuki-Sudo & Kiritani (1991) asserts that all of “these studies on compensatory effect in production and perception suggest that the ISI is a temporal unit in English.” The rationale for this study has originally departed from the rhythmic dichotomy devised by Abercrombie (1967). That is, as in the above study, given only two types of rhythms in world languages, stress-timed and syllable-timed ones, once speakers of a language of a known rhythm type (Korean in the present study as a syllable-timed language), show critical difference in speaking a language of an unidentified rhythm type with a proper rhythm structure, the latter one must be of the other rhythm group. Thus, we have performed production experiments by native speakers of English and Korean to compare their stress-related durational patterns as the previous Japanese study has tried. The experiments will show the difference in duration pattern of ISIs in the sentences in which the number of unstressed syllables between stressed syllables varies. Additionally, we try to clarify the relationship, if any, between the positions of an ISI in a sentence and the duration increase of the ISI. If we obtained
significant results from the experiments we could claim that it is possible for English to be a stress-timed language.

2. Experiments

2.1. Methods

2.1.1 Subjects

Five native speakers of North American English and ten native speakers of Korean participated. The Korean speakers are divided into two groups consisted of five each: proficient speakers of English and non-proficient ones. In the end, however, we got rid of one speaker out of each group since the speech data of them were not appropriate enough to be employed for measurement. The English native speakers included three Americans and one Canadian and they are all English professors of Hankuk University of Foreign Studies (HUFS) or Kyunghee University in Seoul. The two groups of Korean speakers, as noted from the titles, were selected on the basis of their English proficiency. The proficient speakers were undergraduate students of the college of English at Hankuk University of Foreign Studies majoring in translation and interpretation. They all had experiences of staying in overseas countries in which English is a mother tongue or at least an official language. And their stay abroad spanned from 3 to 14 years. The non-proficient speakers, from freshmen to a senior, were also taken from the students of the HUFS. We originally made it a rule that subjects in the second group were not major in English and to reside abroad less than one year before the time of our recording. But, they, at least, have had classes of practical English as an elective course in their freshmen years. All the subjects had normal speech and hearing ability according to their self-report.

2.1.2 Materials

Basically, we employed the same way that Mochizuki-Sudo and Kiritani (1991) used. But, we made whole different sentences and they were distributed and analyzed in a different way. A total of 36 English sentences (12 more in number than the previous Japanese study) were designed for the production experiment. They were allotted first in accordance with the position of the ISI in a sentence: Initial, Medial, and Final. And each position contained three different sets of sentences in which there were four sentences respectively that differed in the number of unstressed syllables intervening between one target stressed syllable and the next. The composition of the three sets in a position was of no desirable correlation among them, but merely of the experimenters’ will to get abundant materials. The reason why we devised four sentences in every single set is coordinate the number of the unstressed syllables in all the sets from zero to three, so that the overall number of syllables in the target ISI ranged from one to four. And an English native speaker who has PhD in linguistics confirmed the grammaticality of the whole sentences.

The sets of sentences are listed in Table 1. As shown beside the set labels, the most important matter was to unify the syllable environments of sentences in a single set. Firstly, the sentences in each set of each position contained two different stressed vowels. What was important was that the first vowels were all short or lax vowels so that they should not affect the length of the target ISIs. Especially, we unified the vowel sorts as the same ones in a single set regardless of the tongue height of them. And the criterion of selecting the unified vowels was that their formant structures and distribution were conspicuously seen on the broad-band spectrograms. Secondly and more importantly, in order not to influence the duration of the ISIs, the consonants following the first target vowels (codas) and preceding the next target vowels (onsets) were all uniformly set as voiceless stops in all sentences, which was affectable to Korean speakers who have no distinction in duration between voiced consonants and the voiceless.

| Set 1. /...æk/ ~/te.../ | Jack testified against his brother. | 1 |
| /...æk/ ~/te.../ | Jackson testified against his brother. | 2 |
| /...æk/ ~/te.../ | Jackson will testify against his brother. | 3 |
| /...æk/ ~/te.../ | Jacksonville will testify against his brother. | 4 |
| Set 2. /...ɪt/ ~/pæ.../ | Kit passed the driving test. | 1 |
| /...ɪt/ ~/pæ.../ | Kitty passed the driving test. | 2 |
| /...ɪt/ ~/pæ.../ | Kitty will pass the driving test. | 3 |
| /...ɪt/ ~/pæ.../ | Kitty would have passed the driving test. | 4 |
| Set 3. /...ɪt/ ~/kɪt/.../ | Poppy kissed her on the cheek. | 1 |
| /...ɪt/ ~/kɪt/.../ | Poppy kissed her on the cheek. | 2 |
| /...ɪt/ ~/kɪt/.../ | Poppy will kiss her on the cheek. | 3 |
| /...ɪt/ ~/kɪt/.../ | Poppy would have kissed her on the cheek. | 4 |

2.1.3 Procedures

Recording has been conducted in a quiet room using Audio Technica Microphone 75 (ATM75) connected to a digital audio tape recorder (SONY RM-D100K) at the sampling rate of 44,000 Hz. And the sentences recorded were digitized to 16-bit mono sound at 16,000 Hz. Labeling and waveform analysis were performed by Praat (version 4204), a speech analysis program which is obtainable online.

The whole sentences were randomly mixed by the experimenter before reading and the same order was used for every subject. All the sentences were stress-marked on each target stressed vowel in the text of the reading sheet to secure the same stress pattern for every speaker. Later, we checked pitch contours from the analysis tool to guarantee that the stresses were properly observed and excluded those which were not.

The subjects were asked to repeat the whole sentences six times in the way that they read from one to thirty-six once in a
row, and did it over and over again to the sixth. Before reading they were given sufficient time for practicing the sentences and two-minute breaks were also given between each reading whenever needed. They could read the text in a comfortable manner and rate. No pauses or mistakes were found in the middle of reading. Giving a pause or making a mistake, they were asked to repeat the relevant sentence from the beginning.

From the total of the six readings we discarded one to obtain five repetitions in all. And we, then, measured the durations of the target ISIs in each sentence. Formant frequency information and waveforms were used to define the onset positions of the vowels. Contrary to Mochizuki-Sudo and Kiritani (1991), in which they also had identified the offsets to measure the length of the stressed vowels, in the present study, since we were not concerned about the compensatory shortening of stressed vowels the offset positions were not included here. In this way, we measured the intervals between the onset of the first stressed vowel and that of the next.

2.2 Results

Two significant results were found from the production experiments by the native English speakers, proficient Korean speakers of English, and non-proficient Korean speakers of English. Firstly, as shown in the Figure 1 the duration of the ISI almost linearly increased as the number of the unstressed syllables did between the target stressed vowels. As Mochizuki-Sudo and Kiritani (1991) noted the duration of the one-syllable ISI was almost the same with one another for all of the three groups. However, the divergence of durations was getting larger and larger as unstressed syllables were added to the target ISI, so that the more unstressed syllables the ISI contained the longer the duration of it became for the non-proficient Korean speakers of English, the less longer for the proficient Korean speakers of English, and the least for the native English speakers. Though the average durations of all sentences in every set of each position showed such an expected result, some heterogeneous outcomes were produced in Set 1 of the Initial and Final positions; the amount of increase in ISI duration was becoming similar or smaller with each other for every group of speakers as the ISIs were getting longer. We first assumed that this had resulted from the difference of position itself of where the target ISI was laid in a sentence. Unfortunately, however, this turned out to have nothing to do with the position of the ISI, which will be looked at later in the second point of this section.

Table 2 also shows another consistent result of the experiment. It clearly says on the right column of the table that the amount of duration increase of an ISI per syllable is much smaller for the English native speakers, a bit larger for the proficient Korean speakers of English, and even larger for the non-proficient Korean speakers of English. This is intuitively confirmed by looking at the increasing numbers from the first group of speakers (NES) to the third (NKS). The absence of data in Set 2 of the Medial position is due to an error in composing a sentence. The word all in the middle of the third sentence in Set 2 of the Medial position necessarily bears a high pitch and long duration of itself since the word is usually considered as a content word accompanied by a pitch accent in a sentence. Consequently, almost every speaker of all the groups pronounced the word with high pitches and relatively long durations, which accordingly led us to exclude the relevant sentence from data measurement. And looking at the Set 3 in the Final position it is thought the native English speakers did not comply successfully with cautions not to insert a pause between him and and in the fourth sentence, which was inferred to cause their values to be a bit higher than that of the proficient Korean speakers of English.

The second point we were looking for was the correlation between the duration of an ISI and the position of it in a sentence. The Japanese study before has claimed that there was found no special relationship between them without giving any specific evidence for it. That is the reason why we arranged the sentences in the positional order. In the first place, we originally hypothesized from the Figure 2 that the relatively longer durations in the Initial and Final positions might be due to an initial strengthening tendency and pre-boundary lengthening effect in English language. Moreover, many other sentences displayed the similar results to the first (Figure 2), which drove us to remain at the primitive conclusion. Especially, for all the three groups of speakers the short Mediial tendency seen in the Figure 2 seemed to be consistently shown in Set 1 of each position. Unfortunately, however, the Set 2 in every position noticeably demonstrated that it was not the case, and the Set 3 at

![Figure 1: Mean durations of ISI of every sentence in each set of each position for English native speakers, proficient Korean speakers of English, and non-proficient Korean speakers of English.](image)

![Table 2: Mean increment in the duration of ISI per syllable.](image)
last almost distorted the first hypothesis into opposite results with the Medial values higher than those of both positions in most sentences. Conclusively, the present study was not successful in proving the pure correlation between the duration of an ISI and its position in a sentence, but as seen from the Figure 2 and other similar cases we cannot necessarily say at the moment that such a relationship does not exist. Further investigations in detail will need conducting.

Figure 2: In Set 1, average ISI durations in different positions of a sentence for Native English speakers.

3. Discussion

Mochizuki-Sudo and Kiritani (1991) has given a first try at one of the traditional controversies both in Phonology and Phonetics by applying the descendent rhythm dichotomy to Japanese native learners of English. In the same context the present study has marked an epoch by doing the job with Korean native speakers. While the previous study has contained both of production and perception experiments, we, at the moment, launched the first step on the production experiment, especially concerning so-called the isochronism in English language rhythm. As the traditional notion of isochronism has not been proved in physical and instrumental ways, the present study also testified against the intuitional concept in experimental way. What is more, though the old-dated invariable idea has been opposed we cannot deny there must be a certain type of distinction in language rhythm. As presented above, on the production level, there were found prominent differences among the native English speakers, proficient Korean speakers of English, and non-proficient Korean speakers of English. The first two groups have produced rather smaller increase in durations of the target ISIs, in the meanwhile the last group of speakers have showed very large divergence in ISI durations. It is not such difficult to relate the effect to the fact that Korean speakers who were not familiar to the English rhythm had problem uttering properly unstressed syllables between stressed syllables, especially as the number of unstressed syllables was increasing. Further investigation needs conducting to see if it’s possible that the fact that Korean speakers, as native speakers of a syllable-timed language, make considerable difference in uttering English language rhythm (let’s say it’s unidentified yet), could be used as critical evidence for the dichotomy of language rhythm between the syllable-timing and stress-timing.

Regarding the variation of ISI locations in a sentence, neither have we found special correlation with the ISI durations. Interestingly, however, quite a few sentences have shown relatively consistent tendency that the duration increase of an ISI became a bit longer at initial and final positions rather than at the medial. Considering it is accepted in English that there is such a tendency for a single utterance to be strengthened at initial and final, precisely at the end of the intonational boundaries, the result of the present study can be thought to have a significant bearing on the issue.

4. References