

# An Acoustic Investigation of the Relationship between Syllable Structure and Foot Duration in English

Masaki TSUDZUKI and Atsunori KAMIYA

## 【要約】

英語のリズムは等時性 (isochrony) によって特徴付けられると考えられている。英語の等時性とは、強勢間の音節数や音節構造が異なろうとも各強勢間すなわちフット間の時間的な距離が一定に近づくことである (Pike, 1945; Abercrombie, 1967)。同化、リエゾン、脱落、強勢移動、弱母音化のような現象が英語母語話者の発話では確認されるが、これらは等時性を得るために生じる現象と考えることができる。また、英語の童謡や詩歌、さらにはアメリカやイギリスの国歌を観察すると、等時性を得やすいように各強勢が配置されていることに気づく。これらのことから英語の等時性は妥当な現象と考えられてきた。しかし実際に英語フットの持続時間を客観的な数値として測定すると、フット間の長さにはかなりの変動がみられることから、多くの研究者はこの時間的差異を根拠に、等時性は主観的な聴覚印象に基づいたものであり、言語の生成面において等時性は存在しないと結論付けている (Shen and Peterson, 1962; O'Connor, 1965; Lea, 1974; Nakatani, O'Connor & Aston, 1981)。また、一方ではフットを構成する音節数が増すと、確かにそのフット全体の持続時間は増加するものの、フットレベルで音節量の補償が働くため各フットに圧縮や伸長がみられることから、不完全ながらもフット間に等時傾

向は存在すると主張する研究者もいる (Uldall, 1971, 1972; Halliday, 1970, 1994)。

上記のように等時性の有無については一致した結論は得られていないが、完全な等時性は存在しないというのが研究者の間の共通認識である。この完全な等時性について神谷 (2010) は、英語母語話者が意識的に等時性を与えた発話においてすら厳密な等時性は観られないこと、そしてその原因の 1 つに英語音素の内在時間が関与することを明確にした。本論文では、さらに英語の音節構造の違いが英語フットの等時性を崩す要因の 1 つであることを明らかにしていく。

## 1. Introduction

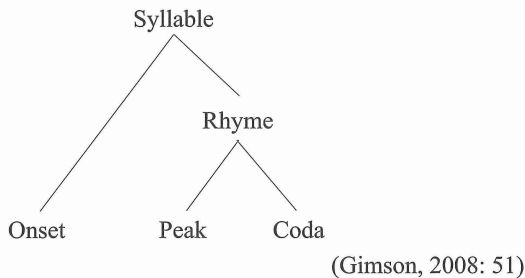
It has been widely claimed that English speakers adjust the duration of intervals between stressed syllables in order to have an isochronal rhythm. This tendency, which is called “*isochrony*,” has been a great concern among *linguists*. Many *researchers* have explored the notion of isochrony in English and have found that English speakers try to maintain fixed intervals between stressed syllables (Huggins, 1975; Fowler, 1977; Schane, 1979). However, with the advent of speech analysis software, many researchers have rejected the validity of isochrony between English feet because they discovered that perfect isochrony is not acoustically possible (Shen & Peterson, 1962; O’Connor, 1965; Lea, 1974; Nakatani, O’Connor & Aston, 1981).

Kamiya (2010) conducted production experiments in order to examine whether it is possible to produce perfect isochrony in English speech. The results revealed that it is not possible to produce evenly spaced feet because the inherent duration of phonemes is not invariable. In this study, production experiments will show whether or not English syllable structure is an additional

factor which impedes isochrony in English speech.

## 2. Syllable Structure

A syllable is a unit consisting of a pivotal vowel or vowel-like phoneme possibly preceded and/or followed by one or more consonant phonemes (Trask, 1996; Gimson, 2008). In metrical phonology, the vowel or vowel-like phoneme is an essential element of the syllable and is known as the “peak,” while the consonant(s) before the peak are called the “onset” and the consonant(s) after the peak are called the “coda” (Lieberman and Prince, 1977). The peak and the coda are referred to as the “rhyme.” This is illustrated in the following diagram:



For example, bit /bɪt/ can be described as CVC and consists of onset, peak, and coda, whereas it /ɪt/ (VC) does not include an onset and bee /bi:/ (CVV) does not include a coda (C means any consonant and V means any vowel or vowel-like phoneme).

Note that the number of syllables in an utterance is the same as the number of peaks in the utterance, but is not related to the number of onsets or codas in the utterance (Roach, 2009).

In the case of pit /pɪt/ and trust /trʌst/, both are composed of onset, peak,

and coda, and are considered to be monosyllabic words because they each include only one vowel. However, their syllable structures are different. The syllable structure of *pit* /pɪt/ is CVC, but *trust* /trʌst/ is CCVCC. In fact, English phonotactics allows the onset to have from zero to three consonants (e.g., *I* (V), *two* (CV), *play* (CCV), *straw* (CCV) and the coda to have from zero to four consonants (e.g., *us* (VC), *aunt* (VCC), *asked* (VCCC), *strengths* (CVCCCC)). These constraints on syllable structure mean that we can identify various syllable structures in English.

However, in other languages, there is little variety in permissible syllable types. For example, in Spanish, the vast majority of syllable types are CV and CVC (Dauer, 1983). Therefore, if we determine the length of a syllable length based on syllable structure alone, it will vary more in English than in Spanish.

As stated above, there is a large variety of permissible syllable structures in English. However, in reality, the distribution among different syllable structures is far from uniform. Dauer (1983) analyzed the frequency of syllable structures in English and found that the majority of stressed syllables have a CVC syllable structure, whereas the majority of unstressed syllables have a CV syllable structure. Moreover, she found that “heavy” and “superheavy” syllables such as CVVC or CVCC in syllable weight occur infrequently. These findings suggest that simple syllable structures tend to occur predominantly in English.

### **3. The Purpose of This Study**

In this study, we will consider onsets of English syllables to determine whether syllable structure is another factor which impedes perfect isochrony in English speech. The method adopted in this study is different from the conventional one because the aim of this study was to investigate whether

English speakers are *artificially* able to produce isochronal feet. If the utterances which English speakers purposely attempted to produce with isochrony still deviate significantly from isochrony, then it can be argued that in English, isochrony of feet can only exist at the psychological level.

#### 4. Subjects

The subjects for this study were three Americans, three British, one Australian, and one Canadian. They were all English teachers at universities and language schools in Japan. All of the subjects did not have any speech or hearing problems. The criterion for selecting subjects was simply that they must all be English native speakers with a higher education. The dialect of English that each speaker speaks was not a factor for consideration.

#### 5. Method and Procedure

The experiment performed here used a rhythmically reiterated nonsense word “teater.” A sequence of the nonsense word “teater” repeated five times (i.e., “/ teater / teater / teater / teater / teater /”) was used as the experimental material. Subjects were instructed to produce the sequence at a natural speed while maintaining a regular rhythm for feet. Next, subjects were asked to change the third “teater” (CVCV) to “streater” (CCCVCV). The purpose of this was to investigate whether different onsets would cause a variation in duration. Both sequences were then compared to determine whether syllable structure is a factor which interferes with isochrony. Additionally, in order to investigate the influence of speaking rate, which might have a strong influence on the actual

duration of “teater” and “streater,” the same experiment was also repeated at a higher speed.

In this study, a reiteration of the two-syllable foot “teater” was adopted as the carrier sequence because two-syllable feet are the most frequently used in English speech. Moreover, this carrier sequence is not affected by the inherent duration of phonemes. Therefore, the carrier sequence, / teater / teater / \_\_\_\_ / teater / teater / is regarded to be an appropriate material in this experiment.

These are the two sequences of nonsense words which were used in this experiment:

1. / teater / teater / teater / teater / teater /
2. / teater / teater / streater / teater / teater /

Considering the influence of constriction duration of closure and final lengthening (Lehiste, 1973; Klatt, 1975, 1976; Lehiste, Olive & Streeter, 1976; Turk & Shattuck-Hufnagel, 2000), the first and last feet were excluded from the analysis. Hence, the duration of the second, third, and fourth feet were measured. The duration of the third foot and the mean duration of the second and fourth feet were subsequently compared using statistical methods.<sup>1</sup>

The utterances were recorded directly onto a personal computer using a microphone (Sony F-V320). After the recording, the duration of each foot was measured in milliseconds using Sugi Speech Analyzer software (made by Animo Limited Corporation). The sound wave images and the sound spectrograms were created in order to locate the foot boundaries. For example, the sound waves, broadband spectrogram, and narrow band spectrogram are shown below for the utterance of “/ teater / teater / streater / teater / teater /” produced by Subject 1 at natural speed.

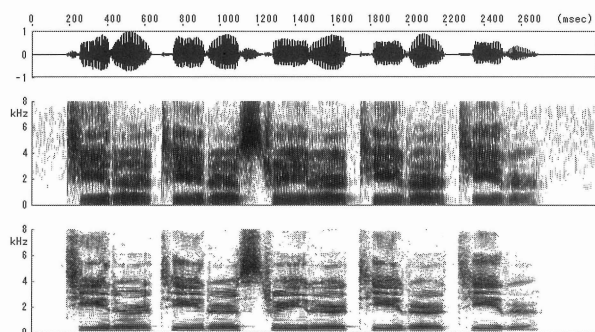


Figure 1: The spectrograms of “/teater/teater/streater/teater/teater/” at Natural Speed

## 6. Results and Discussion

The tables below (Tables 1 to 4) show the duration of the second, third, and fourth foot in each sequence of “/ teater / teater / teater / teater / teater /” and “/ teater / teater / streater / teater / teater /.” The ratio indicates how much the duration of the third foot deviated from the duration of the preceding and following feet. If each foot were produced isochronally, then all of the ratios in the table would be exactly 1. Therefore, the larger the ratio, the greater the divergence from isochronal feet.

When Table 1 and Table 2 are compared, clear differences can be seen in the ratios between the tables. The ratios in Table 1 are much closer to 1 than the ratios in Table 2. This means that the duration of “streater” was produced longer than the duration of “teater” for all of the subjects. To confirm this point, the Mann-Whitney U Test<sup>2</sup> was performed. The result was significant ( $p < 0.01$ ), indicating that there was a significant difference which was caused by the onset.<sup>3</sup> This shows that syllable structure is one of the factors which impedes perfect isochrony in English speech on the physical level at least.

Table 1: “/teater/teater/teater/teater/teater/” at Natural Speed

Natural speed	teater	teater	teater	Ratio
Subject 1	534 ms	514 ms	511 ms	0.98
Subject 2	694 ms	685 ms	696 ms	0.99
Subject 3	664 ms	674 ms	710 ms	0.98
Subject 4	688 ms	703 ms	641 ms	1.06
Subject 5	552 ms	616 ms	618 ms	1.05
Subject 6	914 ms	890 ms	908 ms	0.98
Subject 7	783 ms	886 ms	898 ms	1.05
Subject 8	1027 ms	998 ms	1032 ms	0.97

Table 2: “/teater/teater/streater/teater/teater/” at Natural Speed

Natural speed	teater	streater	teater	Ratio
Subject 1	478 ms	559 ms	529 ms	1.11
Subject 2	885 ms	881 ms	835 ms	1.02
Subject 3	826 ms	919 ms	792 ms	1.14
Subject 4	653 ms	750 ms	699 ms	1.11
Subject 5	891 ms	953 ms	845 ms	1.10
Subject 6	998 ms	1106 ms	1072 ms	1.07
Subject 7	761 ms	908 ms	807 ms	1.16
Subject 8	993 ms	1263 ms	1065 ms	1.23

Table 3: “/teater/teater/teater/teater/teater/” at High Speed

High speed	teater	teater	teater	Ratio
Subject 1	391 ms	400 ms	412 ms	1.00
Subject 2	350 ms	355 ms	352 ms	1.01
Subject 3	522 ms	529 ms	480 ms	1.06
Subject 4	363 ms	349 ms	366 ms	0.96
Subject 5	327 ms	330 ms	318 ms	1.02
Subject 6	626 ms	598 ms	596 ms	0.98
Subject 7	408 ms	410 ms	405 ms	1.01
Subject 8	587 ms	549 ms	538 ms	0.98



Table 4: “/teater/teater/streater/teater/teater/” at High Speed

High speed	teater	streater	teater	Ratio
Subject 1	349 ms	467 ms	386 ms	1.27
Subject 2	309 ms	414 ms	376 ms	1.21
Subject 3	485 ms	565 ms	459 ms	1.20
Subject 4	339 ms	414 ms	341 ms	1.22
Subject 5	292 ms	382 ms	306 ms	1.28
Subject 6	587 ms	701 ms	580 ms	1.20
Subject 7	377 ms	444 ms	351 ms	1.22
Subject 8	515 ms	697 ms	565 ms	1.29

When the test was repeated at high speed, the results were similar. A comparison of Table 3 and Table 4 below shows that the duration of “streater” was longer than the duration of “teater” when subjects produced each sequence at a high speed ( $p < 0.01$ ).

Furthermore, Table 2 and Table 4 were compared to consider how much the speaking rate affects isochrony. The ratio of the duration of the length of “streater” to the average duration of the second and fourth “teater” is clearly much larger at high speed (Table 4) than at natural speed (Table 2). This indicates that isochrony between “streater” and its surrounding “teaters” becomes difficult to attain as the speaking rate increases. The Mann-Whitney U Test showed that there is a significant difference between the ratios at natural speed test and at high speed ( $p < 0.01$ ). This shows that a variation in the speaking rate is one of the factors which impedes perfect isochrony in English speech.

## 7. Conclusion

This study examined English syllable onsets to determine whether syllable

structure is another factor which impedes perfect isochrony in English speech.

In phonology, the term “mora” is defined as a unit which determines syllable weight in some languages (Kubozono & Honma, 2002; Otaka, 2009). Each mora is composed of the peak and the coda.<sup>4</sup> The onset is not part of the mora (Hyman, 1985). For example, there is no difference in the moraic structure (syllable weight) between spit/spit/ (CCVC) and pit/pit/ (CVC) because the number of onsets can be ignored. Moreover, compensatory lengthening, an effect which preserves syllable weight, can be observed only when a syllable-final segment (coda) is deleted, but this never occurs in the case of onsets (Peterson & Lehiste, 1960). Thus, spit/sprt/ (CCVC) is inherently equivalent to pit/pit/ (CVC) from a moraic point of view.

When we adapt the moraic theory to this acoustic study, the duration of “teater” (CVCV) and that of “streater” (CCVCV) are supposed to be the same even if the speaking rate is varied because they are both composed of two moras. However, the results of this study indicated that the actual duration of a foot was affected by the onset(s) of the foot and that isochronal feet were more difficult to be produced at a high speed than at a natural speed.

In conclusion, the present study shows that both syllable structure and speaking rate are important factors which impede perfect isochrony in English speech.

### Notes

- 1 For example, when / teater / teater / streater / teater / teater / was produced, if the length of each word was 403 ms, 389 ms, 424 ms, 402 ms, and 498 ms, respectively, then 424 and 395.5 were compared because  $(389+402)\div 2=395.5$ . Hence, in this example, / streater / was produced 1.061 times longer than surrounding feet because  $424\div 399.5=1.061$ .
- 2 The Mann-Whitney U Test is a non-parametric test for assessing whether or not

two independent samples of observations come from the same distribution.

- 3 We measured the metric feet from the beginning of the onset of a stressed syllable to the onset of the next stressed syllable. Each sequence was divided into feet like this: “/ teater / teater / teater / teater / teater /” and “/ teater / teater / streater / teater / teater /.” However, some researchers start the measurement from one stressed vowel to the next stressed vowel (Allen, 1972). If the measurements had been made based on the latter method, the results might have been different. We have, therefore, reanalyzed the data based on the latter division, i.e. “t / eatert / eatert / eatert / eatert / eater” and “t / eatert / eaterstr / eatert / eatert / eater.” It turns out that the results are the same. The duration of “eaterstr” was produced longer than that of “eatert” for all of the subjects ( $p < 0.01$ ). This confirms that syllable structure is one of the factors which impedes perfect isochrony in English speech.
- 4 Counting the number of moras is used to determine syllable weight in English. According to this notion, a light syllable consists of one mora and a heavy syllable consists of two moras (McCawley, 1968).

### Acknowledgement

This research is supported in part by KAKENHI (Grant-in-Aid for Young Scientists (B) 2011–2013), Grant No. 23720259.

### References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh Univ. Press.
- Allen, G. D. (1972). The location of rhythmic stress beats in English: An experimental study I & II. *Language and Speech*, 15, 72–100, 179–195.
- Dauer, R. M. (1983). Stress-timing and syllable-timing Reanalyzed. *Journal of Phonetics*, 11, 51–62.
- Fowler, C. A. (1977). *Timing control in speech production*. Ph.D. dissertation, University of Connecticut.
- Gimson, A. C. (2008). *Gimson's Pronunciation of English*, 7th ed. Revised by Cruttenden, A., London: Oxford Univ. Press.

- Halliday, M. A. K. (1970). *A Course in Spoken English*. London: Oxford Univ. Press.
- . (1985). *An Introduction to Functional Grammar*, 2nd ed., London: Edward Arnold.
- Huggins, A. W. F. (1975). On isochrony and syntax. In Fant, G. & Tatham, M. A. A. (eds.), *In Auditory analysis and perception of speech*, 455–464, London: Academic Press.
- Hyman, L. M. (1985). *A Theory of Phonological Weight*. Dordrecht, Holland: Foris.
- Kamiya, A. (2010). *An Acoustic Study of Isochronal Feet in English Speech*. Doctoral dissertation, Kwansei Gakuin University.
- Klatt, D. H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research*, 17, 51–63.
- . (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3, 129–140.
- . (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1221.
- Kubozono, H. & Honma, T. (2002). *Onsetsu to Mora* [Syllables and Moras]. Tokyo: Kenkyusya.
- Lea, W. A. (1974). Prosodic aids to speech recognition: A general strategy for prosodically-guided speech understanding. *Univac Report* No. PX10791. St. Paul, Minn.: Sperry Univac, DSD.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Massachusetts: MIT Press.
- . (1973). Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America*, 54(5), 1228–1234.
- . (1975). Some factors affecting the duration of syllabic.
- Lehiste, I., Olive, J. P & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60(5), 1199–1202.
- Lieberman, M. & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2), 249–336.
- McCawley, J. D. (1968). *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.
- Nakatani, L. H., O'Connor, K. D. & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica*, 38, 84–106.
- O'Connor, J. D. (1965). The perception of time intervals. *Progress Report*, 2, 11–15, Phonetics Laboratory, University College, London.

- Otaka, H. (2009). *Phonetics and Phonology of Moras, Feet and Geminate consonants in Japanese*. Maryland: University Press of America.
- Peterson, G. E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693–703.
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor, Mich.: Univ. of Michigan Pr.
- Roach, P. J. (2009). *English Phonetics and Phonology: A Practical Course*, 4th ed. Cambridge: Cambridge Univ. Press.
- Schane, S. A. (1979). The rhythmic nature of English word accentuation. *Language*, 55, 559–602.
- Shen, Y. & Peterson, G. G. (1962). Isochronism in English. *Studies in Linguistics, Occasional papers*, 9, 1–36. Dept. of Anthropology and Linguistics, Univ. of Buffalo., Buffalo, NY.
- Trask, R. L. (1996). *A Dictionary of Phonetics and Phonology*. London: Routledge.
- Tsuzuki, M. (2000). Eigo Purosodi no Anticipation ni tsuite [A Study of Prosodic Anticipation in English]. *English Phonetics*, 3, 255–274.
- Tsuzuki, M. et al. (eds.). (2005). *Eigo Onseigaku Ziten* [The EPSJ Practical Dictionary of English Phonetics]. Tokyo: Seibido.
- Turk, A. E. & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, 28, 398–440.
- Uldall, E. T. (1971). Isochronous Stresses in R.P. *Hammerich-Jakobson-Zwirner*, 205–210.
- . (1972). Relative durations of syllables in two-syllable rhythmic feet in R.P. in connected speech. *Edinburgh University Department of Linguistics, Work in Progress*, 5, 110–111.

(This paper is jointly authored by Masaki TSUDZUKI and Atsunori KAMIYA.)

