



An Analysis of Word Duration in Native Speakers and Japanese Speakers of English

Tomoko Nariai¹, Kazuyo Tanaka¹, Yoshiaki Ito²

¹Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan

²Faculty of Software and Information Science, Iwate Prefectural University, Japan

nariai@slis.tsukuba.ac.jp, ktanaka@slis.tsukuba.ac.jp, y-itoh@iwate-pu.ac.jp

Abstract

An analysis of word duration in English sentences uttered by native speakers of Japanese is made, in which the difference in prosodic patterns between the English and Japanese languages is taken into account. The durations of Japanese speakers are compared with those of English speakers in regard to a percentage distribution of an individual word relative to all words in a sentence. The results of the statistical analysis revealed that nouns and words at the ends of sentences in Japanese speakers were shorter for English speakers. The former result suggests that English speakers put prominence on nouns, whereas Japanese speakers tend not to have the same rhythm as English speakers. The latter result suggests that phrase-final lengthening is insufficient in Japanese speakers.

Index Terms: speech analysis, prominence, Japanese learners of English, second language acquisition

1. Introduction

The revised cumulative guidance record, issued from MEXT (Ministry of Education, Culture, Sports, Science and Technology in Japan), prescribes that English will become a regular course for the fifth and sixth grades at elementary schools in Japan. The early education of English is intended to motivate students to learn to speak English, rather than to read and write in English. As such, clarifying the features of English utterances by Japanese speakers is important in order to facilitate English learning by non-native speakers. This is also important for specialists in computer based educational systems [1, 2].

This paper describes the duration difference between native speakers and Japanese speakers of English on the basis of statistical analysis of sentence speech samples. We presume that irregularities in the English spoken by Japanese typically occur in focus or prominence.

There are a number of studies on the analysis of duration of Japanese speakers of English [1, 2, 3, 4, 5], in which duration of the stressed syllable in a target word was measured. However, lengthening in an accented word in English affects not only the stressed syllable, but also other syllables [6]. In addition, prominence is determined depending on the sentence, therefore, an investigation of only the stressed syllable of a target word is incomplete.

In the present study, irregularities of prominence in Japanese speakers of English are revealed through analysis of word duration with regard to the word class and the word position within a sentence.

In section 2, speech data and concrete procedures for the analysis method are presented. Word durations are analyzed after normalizing the length of each word by the sum of the

lengths of all words contained in a sentence. The word durations are then classified into word classes and position within a sentence. Section 3 shows the analytical results. A discussion of the analytical results is presented in section 4, and the conclusion is given in section 5.

2. Speech samples used in the analysis

2.1. Sample Sentences

One hundred sentences are chosen from MOCHA-TIMIT dataset [7]. The sentence numbers are timit001-030, 211-260 and 441-460. The sentence set consists of 84 declaratives and 16 interrogatives. There are 707 words in total.

2.2. Subjects

2.2.1. Native English speakers

The group of English speakers consisted of 10 subjects, 5 males and 5 females, aged between 20 and 40. Most of the subjects were English teachers living in Japan.

2.2.2. Native Japanese speakers

The group of Japanese speakers consisted of 17 subjects, 9 males and 8 females, aged between 20 and 30. Most of the subjects were undergraduate students. A native English speaker, who was an English teacher in Japan, judged that they were not proficient in English.

2.3. Recording Condition

The subjects were given sufficient time to practice reading the speech materials before recording. Subjects were asked to enunciate clearly and to utter a sentence repeatedly until the speech sample was recorded properly. No other specific instruction for utterances of English was given to subjects.

The 10 English subjects uttered 100 sentences each.

A group of 9 Japanese subjects uttered 50 sentences each, sentence numbers of which are timit001-030 and 211-230. A second group of 8 Japanese subjects uttered the remaining 50 sentences, i.e., timit231-260, and 441-460.

2.4. Method of Analysis

Each sentence utterance is sampled at the rate of 16 kHz and quantized into 16 bits. The acoustic feature extraction is conducted by WaveSurfer. Extracted word durations of individual sentences are segmented into word sequences, where word boundary is determined by observing the waveform and the spectrogram patterns. The word with cases such as vowel epenthesis or reduction, or a phonological clitic of monosyllabic word is not deleted as long as it does not interfere in word boundary detection.

Statistical significance of the difference in word duration between the two groups can be evaluated on the basis of the criterion used in statistical pattern recognition, that is, the ratio of between-group variance to within-group variance, known as Fisher's ratio in linear discriminant analysis. We denote this ratio as R. If R is large, a considerable difference exists in the sample distribution of two groups. The procedure for calculating R is as follows.

(1) Each word duration is normalized by the sum of word durations in a sentence. That is,

$$x_j(i)' = \frac{\sum_{i=1}^L x_j(i)}{L}$$

$$y_j(i)' = \frac{\sum_{i=1}^L y_j(i)}{L}$$

$x_j(i)$: duration of word i uttered by an English speaker j

$y_j(i)$: duration of word i uttered by a Japanese speaker j

L : number of word contained in a sentence

(2) Mean values and variances of $x_j(i)'$ and $y_j(i)'$ for the English speaker group and the Japanese speaker group are calculated as follows:

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j(i)'$$

$$\bar{y} = \frac{1}{M} \sum_{j=1}^M y_j(i)'$$

$$\sigma_x = \frac{1}{N} \sum_{j=1}^N (x_j(i)' - \bar{x})^2$$

$$\sigma_y = \frac{1}{M} \sum_{j=1}^M (y_j(i)' - \bar{y})^2$$

N: number of English speakers

M: number of Japanese speakers

(3) Thus R is obtained as follows:

$$R = \frac{(\bar{x} - \bar{y})^2}{(\sigma_x + \sigma_y)}$$

2.5. Class

2.5.1. Word class

Individual words are classified as content words or function words. Content words are further classified as a noun (core of a noun phrase), verb (core of a verb phrase, including the present progressive form and passive verb), adjective (plays the role of an adjective in a sentence), and adverb (plays the role of an adverb in a sentence).

Function words are also further classified as conjunction / preposition, be/auxiliary verb / do, article, pronoun (including it's) and interrogative / negative (including don't).

2.5.2. Categories of word position

Words are divided into the following three positions: words at the beginning of sentences; words within sentences that do not begin or end sentences; and words at the ends of sentences. The words at the ends of subordinate sentences, which are defined as second end in the following section, are included in words at the ends of sentences.

2.5.3. Second end

The second end is defined as follows.

Table 1: Results for content words.

	noun	adj	verb	adv
number of word	189	135	106	21
$R > 0.2$	113	70	54	13
$ntv > jpe$	102 (90%)	43 (61%)	24	4
$ntv < jpe$	11	27	30 (56%)	9 (69%)

(i) If a main sentence has a prepositional phrase or adverb phrase separated by a comma, the word before the comma is defined as a second end.

(ii) If a sentence is such a complex sentence as to have a second verb in the following subordinate conjunctive phrase or coordinative conjunctive phrase, the word before the conjunction is defined as a second end.

3. RESULTS

The mean and variances of x and y , (\bar{x}, σ_x) and (\bar{y}, σ_y) were obtained from samples of each group. Then R values, from R_1 to R_{707} , were obtained from 707 words. Here, "ntv>jpe" indicates that the word duration for the group of English speakers "ntv" is longer than that for the group of Japanese speakers "jpe", and "ntv<jpe" indicates that the word duration for Japanese speakers is longer than that for English speakers. In the present study, only cases of significant difference between the two groups were counted; specifically, inequality, "ntv>jpe" or "ntv<jpe", holds only for cases that satisfy $R > 0.2$.

3.1. Ratio of Word Class

3.1.1. Content word

Table 1 shows a summary of the results for content words. There were 451 content words in 100 sentences: 189 nouns, represented by 'noun'; 135 adjectives, 'adj'; 106 verbs, 'verb'; and 21 adverbs, 'adv'. For each word class, the word durations in cases of "ntv>jpe" and "ntv<jpe" are determined.

Out of the 189 nouns, 113 satisfy $R > 0.2$, 90% of which satisfy "ntv>jpe." Out of the 135 adjectives, 70 satisfy $R > 0.2$, 61% of which satisfy "ntv>jpe." In contrast, out of the 106 verbs, 54 satisfy $R > 0.2$, 56% of which satisfy "ntv<jpe." Out of the 21 adverbs, 13 satisfy $R > 0.2$, and nine of the adverbs satisfy "ntv<jpe." These suggest that the probability of "ntv>jpe" for nouns is the highest among content words. This means that the durations of nouns for Japanese speakers tend to be shorter than those for English speakers.

In Figure 1, the mean duration and standard deviation of words in *timit259*, "Do they allow atheists in church," are plotted for English and Japanese speakers. The words for "ntv>jpe" are *atheists*_{(5.90)(noun)} and *church*_{(1.19)(noun)}, where the R value for each word is given in the first parenthesis and the word class is given in the second parenthesis. The word for "ntv<jpe" is *allow*_{(0.35)(verb)}. These data suggest that the durations of nouns for English speakers tend to be longer than those for Japanese speakers. Therefore, it can be said that irregularities in prominence by Japanese speakers tend to occur on nouns, compared with English speakers.

3.1.2. Function word

Table 2 shows the results for function words. There are 256 function words in 100 sentences: 12 interrogatives/negatives,

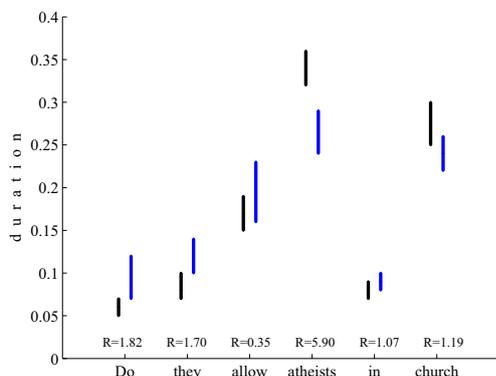


Figure 1: Durations of content words for a group of English speakers (left-side bar) and a group of Japanese speakers (right-side bar)

Table 2: Results for function words.

	inng	cnj	be	art	prn
number of word	12	84	37	74	49
$R > 0.2$	4	67	24	47	29
$ntv > jpe$	3 (75%)	1	4	1	2
$ntv < jpe$	1	66 (99%)	20 (83%)	46 (98%)	27 (93%)

'inng'; 84 conjunctions/prepositions, 'cnj'; 37 be/auxiliary verb/do, 'be'; 74 articles, 'art'; and 49 pronouns, 'prn'.

For the 12 interrogatives/negatives, 4 words satisfy $R > 0.2$, and 75% of which satisfies "ntv > jpe". On the contrary, for $R > 0.2$, 99% of the 67 conjunctions/prepositions, 83% of the 24 be/auxiliary verb/do, 98% of the 47 articles, and 93% of the 29 pronouns satisfy "ntv < jpe."

Figure 2 shows the result of words in timit228, "How good is your endurance?," where *How*_{(0.46)(interrogative)} satisfies "ntv > jpe," however, *is*_{(0.98)(be)} and *your*_{(1.12)(pronoun)} satisfy "ntv < jpe."

The above results suggest that the durations of the majority of function words for Japanese speakers are longer than those for English speakers. This can be said that prominence is misplaced on most function words uttered by Japanese speakers.

3.2. Ratio of Word Position

Table 3 shows the results for words at the beginning of and within sentences. There are 100 words at the beginning of sentences, represented by 'beginning', that include 38 content words, 'C', and 62 function words, 'F.' There are 497 words within sentences, 'within,' that include 308 content words and 189 function words. Table 4 shows the results of 100 words at the ends of sentences, and 10 second ends, 'secen.' The words at the ends of sentences are also classified as 84 words appearing in declarative sentences, 'dec,' and 16 words appearing in interrogative sentences, 'inter.'

For function words at the beginning of and within sentences, the results suggest that the durations of the majority of words for Japanese speakers are longer than those for English speakers.

For content words at the beginning of and within sentences, the results indicate that the number of words for which "ntv > jpe" holds is slightly more than the number of words for which "ntv < jpe" holds.

In contrast, for content words at the ends of declarative sentences and of interrogative sentences, and at second ends, the results suggest that the durations of the majority of the end words

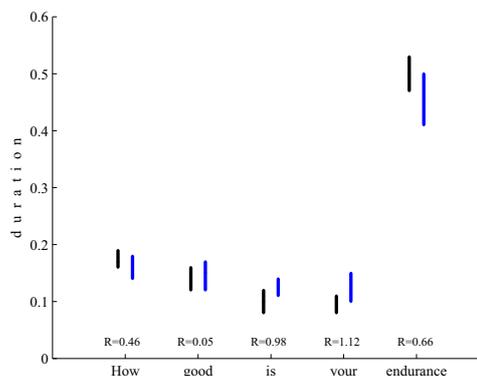


Figure 2: Durations of interrogative "how" for English speakers and Japanese speakers

Table 3: Results for words at the beginning of and within sentences.

	beginning		within	
	C	F	C	F
number of word	38	62	308	189
$R > 0.2$	18	36	143	132
$ntv > jpe$	11 (61%)	4	84 (59%)	4
$ntv < jpe$	7	32 (89%)	59	128 (97%)

Table 4: Results for end words of interrogative sentences and declarative sentences, and second ends.

	inter	dec		secen	
	C	C	F	C	F
number of word	16	80	4	9	1
$R > 0.2$	12	56	3	5	0
$ntv > jpe$	12 (100%)	53 (94%)	2 (67%)	4 (80%)	
$ntv < jpe$		3	1	1	

for English speakers are longer than those for Japanese speakers.

Next, we consider examples in which the same word appears within a sentence and at the end of a sentence. Figure 3 shows the result for timit019, "Aluminum cutlery can often be flimsy," where the auxiliary verb *can*_(1.57) exists in the sentence. Here, *can* satisfies "ntv < jpe." Figure 4 shows the result for timit013, "Swing your arm as high as you can," where the auxiliary verb *can*_(0.83) ends the sentence. Here, *can* satisfies "ntv > jpe." Thus, the duration of the auxiliary verb *can* satisfies "ntv < jpe" when the word appears within a sentence, but satisfies "ntv > jpe" when the word appears at the end of sentence.

In conclusion, the results suggest that, for English speakers, prominence lies on words at the ends of sentences, and that irregularities in prominence by Japanese speakers occur for words at the ends of sentences. The results also suggest that prominence is misplaced on function words at the beginning of and within sentences for Japanese speakers.

4. Discussion

The results depending on the word class in section 3.1 revealed that irregularities in prominence occur on nouns, and that prominence is misplaced on most function words in Japanese English.

Our results confirm the results of previous studies [8, 9], indicating that Japanese speakers of English tended to utter an

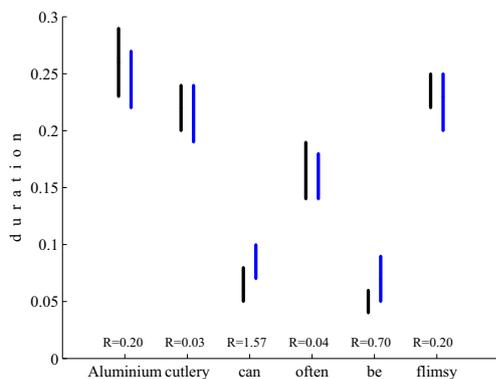


Figure 3: Durations of the auxiliary verb 'can' that appears within a sentence

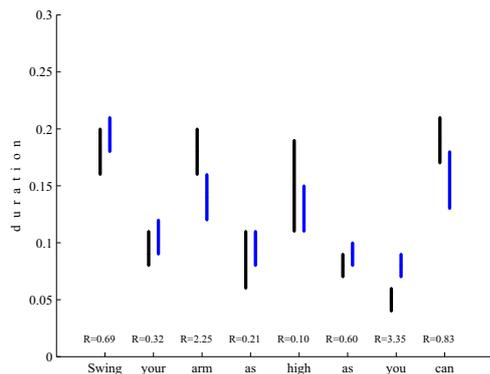


Figure 4: Durations of the auxiliary verb 'can' that appears at the end of a sentence

important word (i.e., content word) non-emphatically. Also, in English, lexically stressed syllables are longer in duration than unstressed syllables. Therefore, our results correspond to the knowledge gained from previous studies that in Japanese speakers of English, there is no significant duration contrast between stressed and unstressed syllables [3, 4]. In addition, the result of noun lengthening by English speakers extends to the knowledge used in a prosody learning system [1, 2], in which the stressed syllable in sentences in a learner's utterance is detected. Our results indicate that the word with the most stressed syllable is a noun, and therefore, that English speakers put prominence on nouns, whereas Japanese speakers of English tend not to have the same rhythm as English speakers.

The results depending on the word position in section 3.2 revealed that irregularities of prominence occur on words at the ends of sentences in Japanese English. Words at the ends of sentences are the most appropriate word of prominence in English [10].

A phrase-final word, in a precise sense, corresponds to second end in this study, but is just as valid for a number of cases of sentence end in this study. Phrase-final lengthening is usually considered to occur as a cross-linguistics feature in every spoken language (e.g., for English [11, 12]; for Japanese [13, 14]). Also, for Japanese speakers of English, the previous study by Ueyama [5] suggested that there is a possible positive transfer of Japanese at the end of a prosodic unit. The difference in results between her study and the present study can be explained by the following two factors. One is attributed to the different scope of the measured duration. Ueyama examined the syllable duration, while we examined the word duration normalized by the length of individual sentence utterances. Second, she gave a detailed analysis of several samples uttered by several speakers,

while we statistically analyzed a number of sentences uttered by a number of speakers. Considering the two studies together, we conclude that the effect of prominence in English spreads not only to the length of the stressed syllable, but also to that of other syllables in the word; for English speakers, the duration at the end of sentence is affected by the combination of the effect of irregularities and the spreading features of phrase-final lengthening effect. In contrast, for Japanese speakers, the length at the end of sentence is affected by only the effect of phrase-final lengthening. In consequence, phrase-final lengthening is insufficient in Japanese speakers of English.

5. Conclusion

Statistical analysis of word duration between native speakers and Japanese speakers of English was conducted. The results of the statistical analysis revealed that irregularities of prominence occur most often for nouns, and that prominence is misplaced on most function words by Japanese speakers of English. The results also revealed that irregularities of prominence occurred most often for words at the ends of sentences.

This research is supported in part by a Grand-in-Aid for Scientific Research, Project No.22500145, from the Japan Society for the Promotion of Science.

6. References

- [1] N. Minematsu, S. Kobayashi, K. Hirose and D. Erickson, "Acoustic modeling of sentence stress using differential features between syllables for English rhythm learning system development," *Proc. Int. Conf. Spoken Language Processing*, pp. 529–532 (2002).
- [2] K. Imoto, Y. Tsubota, A. Raux, T. Kawahara, M. Dantsuji, "Modeling and automatic detection of English sentence stress for computer-assisted English prosody learning system," *Proc. Int. Conf. Spoken Language Processing*, pp. 749–752 (2002).
- [3] T. Tarui, "Rhythm timing in Japanese English," *Proc. Int. Conf. Spoken Language Processing*, pp. 592–595 (2000).
- [4] S. Nakamura, H. Kato, Y. Sagisaka, "Effects of mora-timing in English rhythm control by Japanese learners," *Proc. of Inter-Speech*, pp. 1539–1542, (2009).
- [5] M. Ueyama, "Phrase-final lengthening and stress-timed shortening in the speech of native speakers and Japanese learners of English," *Proc. Int. Conf. Spoken Language Processing*, 2, pp. 610–613 (1996).
- [6] A.E. Turk and J.R. Sawusch, "The domain of accentual lengthening in American English," *J. Phonetics*, 25, 25–41 (1997).
- [7] <http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html>
- [8] M. Sugito, *Nihonjin no eigo* (Izumishoin, Tokyo, 1996). (in Japanese)
- [9] K. Watanabe, *Eigo no rhythm, intonation no shido*, (Taishukan-shoten, Tokyo, 1994). (in Japanese)
- [10] R. Quirk, S. Greenbaum, G. Leech, J. Svartvik, *A Comprehensive grammar of the English language* (Longman, London, 1985), pp. 1356–1357.
- [11] D.K. Oller, "The effect of position in utterance on speech segment duration in English," *J. Acoust. Soc. Am.*, 54, 235–247 (1973).
- [12] A.E. Turk, S. Shattuck-Hufnagel, "Multiple targets of phrase-final lengthening in American English words," *J. Phonetics*, 35, 445–472 (2007).
- [13] N. Kaiki, Y. Sagisaka, "The control of segmental duration in speech synthesis using statistical methods," In *Speech Perception, Production and Linguistic Structure*, Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisaka Eds. (Ohmsha, Tokyo, 1992).
- [14] N. Campbell, "Segmental elasticity and timing in Japanese speech," In *Speech Perception, Production and Linguistic Structure*, Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisaka Eds. (Ohmsha, Tokyo, 1992).