

Comparative Analysis of Intensity between Native Speakers and Japanese Speakers of English

Tomoko Nariai¹, Kazuyo Tanaka² and Tatsuya Kawahara³

¹ Department of Japanese Language Arts, Ibaraki Women's Junior College, Japan

² Faculty of Library, Information and Media Science, University of Tsukuba, Japan

³ Academic Center for Computing and Media Studies, Kyoto University, Japan

nariai@slis.tsukuba.ac.jp, ktanaka@slis.tsukuba.ac.jp, kawahara@i.kyoto-u.ac.jp

Abstract

Intensity has been reported as a reliable acoustical correlate of stress accent for English language, but not of pitch accent for Japanese language. This difference between English and Japanese languages is presumed to shape the characteristics of intensity in English spoken by Japanese (*Japanese English*, henceforth). Based on this presumption, the intensity of words in sentence utterances for Japanese English is compared to that for native speakers' English (*native English*, henceforth). Statistical analysis shows that nouns for Japanese English are produced with less intensity, whereas most function words are with more intensity than those for native English. A correlation is recognized between the above results and the proficiency in Japanese English.

Index Terms: power patterns, amplitude, prominence, sentence stress, second language production

1. Introduction

Research in language production agrees that prosodic features play a pivotal role in human communication [1]. Therefore, learning prosodic patterns is essential to acquire second language pronunciation.

There has been a significant increase in studying the characteristics in Japanese English. However, most of the research to date has focused on the segmental level [2, 3], and research into the prosodic level [4] remains relatively sparse. In addition, few studies on Japanese English have dealt with prosodic features, especially intensity, quantitatively.

In this paper, we investigate the intensity that associated with duration, as an acoustic parameterization of loudness that is evaluated with a psychological element. Intensity of words in sentence utterances between native English and Japanese English is compared, and then divided with regard to word class.

In Section 2, the sample data and the analysis method are presented. Section 3 shows the analytical results. The results after the classification of Japanese subjects according to English proficiency are also shown. A dis-

cussion is presented in Section 4, and the conclusion is given in Section 5.

2. Speech samples used in the analysis

2.1. Sample Sentences

The sentence text set of this speech dataset is the same as that of the TIMIT dataset [5]. In the analysis, 100 sentences are chosen; the sentence numbers are timit001-030, 211-260, and 441-460. There are 707 words in total.

2.2. Subjects

2.2.1. English speakers

This group of native English speakers consists of 10 subjects, five males and five females, aged between 20 and 40 years. Most of the subjects were English teachers living in Japan, and were from the United Kingdom, Canada, New Zealand, Australia, and the United States.

2.2.2. Japanese speakers

The group of native Japanese speakers consists of 24 subjects, 12 males and 12 females, aged between 20 and 30 years.

2.3. Recording Conditions

The subjects were given sufficient time to practice reading the speech materials before recording. Subjects were asked to enunciate clearly and to utter the 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 12 Japanese subjects uttered 50 sentences each, the sentence numbers of which are timit 001-030 and 211-230. A second group of 12 Japanese subjects uttered the remaining 50 sentences.

Each sentence utterance was sampled at the rate of 48 kHz and quantized into 16 bits.

2.4. Analysis Method

The intensity value of a word is calculated from the power pattern of each utterance, using WaveSurfer / Snack Sound Toolkit [6]. The power pattern is extracted from the speech wave with a Hamming window of length 20 ms, at a frame interval of 10 ms, and converted into dB. The maximum of the power value is 80 dB, and the minimum is 0 dB. The power value sequence is manually segmented into a word sequence.

The intensity of word is defined as

Intensity of word (i) = Sum of the power values for the section of word i.

The intensity values of individual words are then normalized by the average of those values contained in the sentence. The intensity of word i is represented by the following.

$$x_j(i)' = x_j(i) / \bar{x}_j \quad (1)$$

$$y_j(i)' = y_j(i) / \bar{y}_j \quad (2)$$

where,

$$\bar{x}_j = \sum_{i=1}^L x_j(i) / L$$

$$\bar{y}_j = \sum_{i=1}^L y_j(i) / L$$

$x_j(i)$: the intensity of word i uttered by English speaker j

$y_j(i)$: the intensity of word i uttered by Japanese speaker j

L : number of words in the corresponding sentence

In this study, an index, denoted by R , is used for representing the difference between the two groups. This R is conventionally used in statistical pattern recognition and is the ratio of the between-group variance to the within-group variance, known as Fisher's ratio in linear discriminant analysis. A large R indicates a significant difference between the two groups. R is calculated as follows.

$$R(i) = \frac{(\bar{x}(i) - \bar{y}(i))^2}{(\sigma_x(i) + \sigma_y(i))} \quad (3)$$

where,

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

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

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

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

N : number of English speakers

M : number of Japanese speakers

R values are calculated for individual words, then, we obtain R_1 to R_{707} .

In the analysis, “ntv>jpe” indicates that a word for English subjects “ntv” is with more intensity than that for Japanese subjects “jpe,” and “ntv<jpe” indicates the reverse. In the present study, only cases of significant difference between the two groups are counted; specifically, inequality, “ntv>jpe” or “ntv<jpe” holds only for cases that satisfy $R > 0.2$.

2.5. Classification

Individual words are classified into content words or function words.

Content words are further classified into nouns (core of noun phrases), verbs (core of verb phrases), adjectives (play the role of adjective in sentences), and adverbs (play the role of adverb in sentences).

Function words are also further classified into five groups of /interrogative, negative/, /conjunction, preposition/, /be, auxiliary verb, do/, article and pronoun.

3. Results

3.1. Word Class

3.1.1. Content Word

Table 1 shows the results for content words. The columns in the table indicate noun, represented by ‘noun,’ adjective, ‘adj,’ verb, ‘verb,’ and adverb, ‘adv.’ The second row from the top indicates the number of words classified into the word class. The third row indicates the number of words that satisfy $R > 0.2$. The fourth and the fifth rows indicate the number of words that satisfy “ntv>jpe” and “ntv<jpe,” respectively.

From the table, we can see that noun amounts to 190, adjective to 134, verb to 106, and adverb to 22. For each word class, the number of words satisfying “ntv>jpe” and “ntv<jpe” is counted.

Out of 190 nouns, 111 satisfy $R > 0.2$, 87% of which satisfy “ntv>jpe.” Out of 134 adjectives, 61 satisfy $R > 0.2$, 59% of which satisfy “ntv>jpe.” In contrast, out of 106 verbs, 51 satisfy $R > 0.2$, 53% of which satisfy “ntv<jpe.” Out of 22 adverbs, 11 satisfy $R > 0.2$, 64% of which satisfy “ntv<jpe.”

The above results suggest that more than half the nouns and adjectives satisfy “ntv>jpe,” and more than half the verbs and adverbs satisfy “ntv<jpe.” We can identify the following characteristics for content words in Japanese English.

Figure 1 shows the mean and standard deviation of intensity for words in timit242, “Do they allow atheists

Table 1: Results for content words.

| | noun | adj | verb | adv |
|----------------|-------------|-------------|-------------|------------|
| number of word | 190 | 134 | 106 | 22 |
| $R > 0.2$ | 111 | 61 | 51 | 11 |
| $ntv > jpe$ | 97 (87%) | 36 (59%) | 24 | 4 |
| $ntv < jpe$ | 14 | 25 | 27 (53%) | 7 (64%) |

Table 2: Results for function words.

| | inng | cnj | be | art | prn |
|----------------|-------------|--------------|-------------|-------------|-------------|
| number of word | 11 | 84 | 37 | 74 | 49 |
| $R > 0.2$ | 4 | 71 | 25 | 48 | 27 |
| $ntv > jpe$ | 4 (100%) | | 4 | 3 | 5 |
| $ntv < jpe$ | | 71 (100%) | 21 (84%) | 45 (94%) | 22 (82%) |

in church?” The words that satisfy “ $ntv > jpe$ ” are nouns, *atheists*_(3.8) and *church*_(0.97), where the R is given in the parenthesis. In contrast, the verb, *allow*_(0.39), satisfies “ $ntv < jpe$.”

Figure 2 shows the result of words in timit451, “*The toddler found a clamshell near the camp site.*” The nouns and adjective, *toddler*_(0.26), *clamshell*_(0.95), *camp*_(0.73) and *site*_(0.81), satisfy “ $ntv > jpe$.” In contrast, the verb, *found*_(0.41), satisfies “ $ntv < jpe$.”

Also, the probability of “ $ntv > jpe$ ” for nouns is the highest among content words. This means that nouns for Japanese English are produced with less intensity than for native English.

3.1.2. Function Word

Table 2 shows the results for function words. Function words amount to 255 words: 11 /interrogative, negative/, ‘int, ng’; 84 /conjunction, preposition/, ‘cnj, prp’; 37 /be, auxiliary verb, do/, ‘be’; 74 articles, ‘art’; and 49 pronouns, ‘prn.’

Out of 11 /interrogative, negative/, 4 satisfy $R > 0.2$, all words satisfy “ $ntv > jpe$.” On the contrary, over 80% of /conjunction, preposition/, /be, auxiliary verb, do/, article, and pronoun satisfy “ $ntv < jpe$.” From these results, we can identify the following characteristics for function words in Japanese English.

Figure 3 shows the result of words in timit221, “*How permanent are their records,*” where the interrogative, *How*_(0.39), satisfies “ $ntv > jpe$.” In contrast, the be, *are*_(4.64), and the pronoun, *their*_(1.26), satisfy “ $ntv < jpe$.”

Figure 4 shows the result of words in timit242, “*The courier was a dwarf,*” where the articles, *The*_(0.73) and *a*_(0.21), and the be, *was*_(2.97), satisfy “ $ntv < jpe$.”

The above results suggest that most function words for Japanese English are produced with more intensity

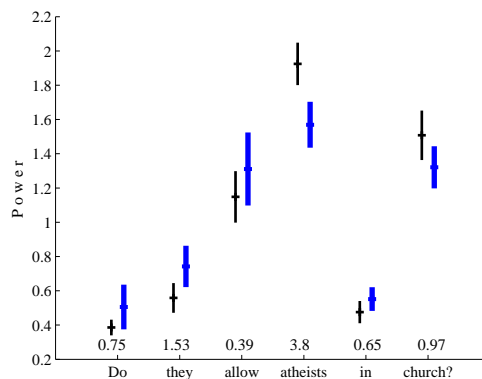


Figure 1: The intensity of words in timit259 for English subjects (left-side thin bar) and Japanese subjects (right-side bold bar). Each bar indicates the range from (mean - SD) to (mean + SD).

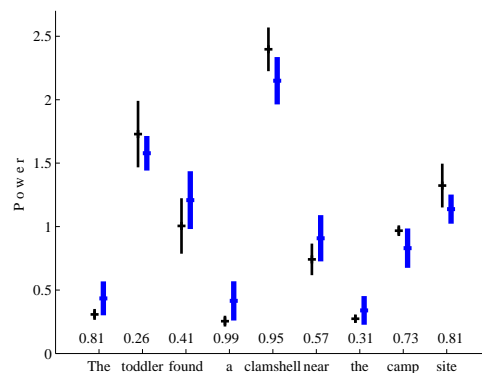


Figure 2: The intensity of words in timit451 for English subjects and Japanese subjects.

than those for native English.

3.2. Classification of Japanese subjects' proficiencies

The results in subsections 3.1.1 and 3.1.2 revealed that nouns for Japanese English were produced with less intensity, whereas most function words for Japanese English were with more intensity than those for native English. In this subsection, these results are considered according to the correlation of the English proficiency of Japanese subjects.

Japanese subjects are classified according to English proficiency. A native speaker of English, who was an English teacher in Japan, listened to the utterances of all subjects, and judged their proficiency levels. The gauge for determining proficiency level is a ten-point scale, ranging from “fully Japanese spoken English” to “natively spoken English.” Each of the 24 subjects is assigned a level from two to eight. 17 of the subjects are classified into the levels of five and below and 7 into the levels of six or greater.

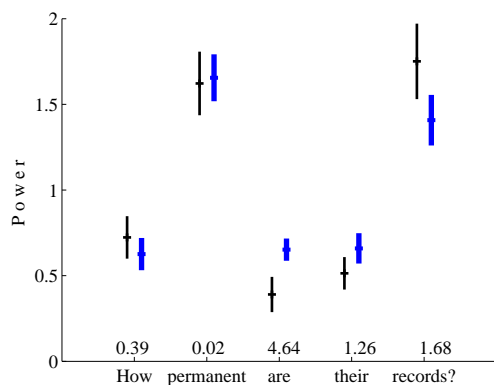


Figure 3: The intensity of words in timt221 for English subjects and Japanese subjects.

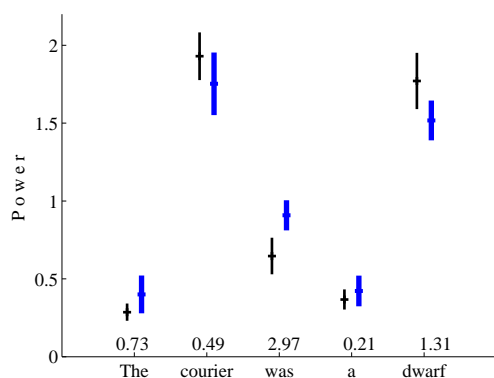


Figure 4: The intensity of words in timt242 for English subjects and Japanese subjects.

Table 3 shows the results for 190 nouns, ‘noun,’ and 255 function words, ‘function.’ The levels of five and below are referred to as ‘intermediate,’ and the levels of six or greater as ‘advanced.’

For nouns, the results suggest that the probability of satisfying “ $ntv > jpe$ ” for the intermediate prosody level is higher than that for the advanced prosody level. Also, the results for function words suggest that the probability of satisfying “ $ntv < jpe$ ” for the intermediate prosody level is higher than that for the advanced prosody level.

These data suggest that Japanese English has the characteristics such as the results in subsections 3.1.1 and 3.1.2.

4. Discussions

The analyses in subsections 3.1.1 and 3.1.2 found that nouns for Japanese English were produced with less intensity, whereas most function words for Japanese English were with more intensity than those for native English.

Our findings confirm the results of previous studies [7, 8], indicating that Japanese English tends to utter an important word (i.e., content word) non-emphatically.

Table 3: Results of the classification of Japanese subjects.

| | advanced | | intermediate | |
|----------------|-------------|--------------|--------------|--------------|
| | noun | function | noun | function |
| number of word | 190 | 255 | 190 | 255 |
| $R > 0.2$ | 119 | 168 | 116 | 185 |
| $ntv > jpe$ | 91 (77%) | 25 | 102 (88%) | 17 |
| $ntv < jpe$ | 28 | 143 (85%) | 14 | 168 (91%) |

In addition, our findings add knowledge such that an important word that is designated in the previous studies is a noun, and that Japanese English applies much prominence to function words, that cause nouns non-prominent.

Also, our findings can extend to the knowledge used in a prosody learning system [9], in which the stressed syllable in sentences in a learner’s utterance is detected. Our results indicate that the stressed syllable is a noun, and that Japanese English does not apply prominence on the noun.

5. Conclusions

We described the characteristics of Japanese English sentences that occurred in intensity in relation to the word class. The statistical analyses revealed that nouns for Japanese English were produced with less intensity, whereas most function words for Japanese English were with more intensity than those for native English. The knowledge gained by the analysis can be used in English education and computer-based educational systems.

6. References

- [1] J. Hirschberg, “Communication and Prosody: Functional Aspects of Prosody,” *Speech Communication*, **36**, 31–43 (2002).
- [2] K. Hattori, P. Iverson, “English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy,” *J. Acoust. Soc. Am.*, **125**, Issue 1, 469–479 (2009).
- [3] K. Saito, “Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English,” *RELC Journal*, **42**, 363–378 (2011).
- [4] M. E. Beckman, J. B. Pierrehumbert, “Intonational Structure in Japanese and English,” *Phonology Yearbook*, **3**, 255–310 (1986).
- [5] <http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html>
- [6] K. Sjlinder, J. Beskow, “Wavesurfer - an open source speech tool,” *Proc. ICSLP*, **4**, pp. 464–467, 2000.
- [7] K. Watanabe, *Eigo no rhythm, intonation no shido*, (Taishukan-shoten, Tokyo, 1994). (in Japanese)
- [8] M. Sugito, *Nihonjin no eigo* (Izumishoin, Tokyo, 1996). (in Japanese)
- [9] 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. IC-SLP*, pp. 749–752, 2002.