A Study of Pitch Patterns of Sentence Utterances by Japanese Speakers of English in Comparison with Native Speakers of English

Tomoko Nariai[1], *Kazuyo Tanaka*[1]

[1]Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
nariai@slis.tsukuba.ac.jp, ktanaka@slis.tsukuba.ac.jp

Abstract

This paper describes statistical analyses for identifying certain inherent ambiguities on pitch patterns of sentence utterances in English spoken by Japanese (Japanese English, henceforth). Statistical significance of pitch pattern differences between Japanese English and native English speakers is evaluated depending on the word position in a sentence and the word class, such as content word and function word. Results suggest that in Japanese English, sentences have lower pitch at the beginning and higher pitch at the end than sentences uttered by English speakers. Also, pitch ranges in sentences in Japanese English are narrower than those for English speakers. These indicate that intonation pattern in Japanese English is rather flat. Additionally, the results suggest that function words in Japanese English have higher pitch than English speakers.

Index Terms: Japanese learners of English, second language learning, speech analysis, prosody

1. Introduction

Japanese-speaking English often involves somewhat corrupt form, which is known as *Japanese English*. Certain features of Japanese English have been elucidated in foregoing studies [1]. However, few studies of Japanese English have shown quantitative evidence of features in sentence utterance structure, i.e. sentence pitch.

Pitch patterns in Japanese English have revealed as causes of the unnaturalness [2]. Previous study [3] investigated sentence pitch in Japanese English statistically. However, it only examined maximal and minimal frequencies in a pitch pattern of each Japanese English sentence. Therefore subtilizing sentence pitch admits of further research. The present study analyzes pitch patterns of individual words in sentences.

The words in sentence pitch by a group of Japanese English are statistically compared to those by a group of native English speakers, referred to in this article as *Native English*. Concrete procedure and analysis results are presented in the following sections.

2. Analysis Method

2.1. Outline of Analysis Method

Each sample utterance is analyzed by WaveSurfer from KTH, to extract fundamental frequency patterns, and is segmented into words. (Henceforth, the fundamental frequency pattern is considered as equivalent to the pitch pattern.) Then the value of pitch peak, denoted by $peak(i)$, in individual word i is estimated. The values is defined as:

$peak(i)$= *maximal fundamental frequency of word i*

Then words are classified by (a) sentence position and (b) word class. In both cases, following two parameters are examined: *Ave* and *R*, described in section 2.2.1 and 2.2.2. To examine pitch range in a sentence, following three parameters are examined for each sentence: *Max*, *Min* and *D- range*, described in section 2.2.3.

2.2. Statistical Measure Used in the Analysis

2.2.1. Ave value

Each word is normalized by the average of the words contained in the corresponding utterance of a sentence, as in equation (1).

$$\begin{aligned} x(i)' &= x(i)/x_0 \\ y(i)' &= y(i)/y_0 \end{aligned} \tag{1}$$

where

$$x_0 = \sum_{i=1}^{L} x(i) \Big/ L, \quad y_0 = \sum_{i=1}^{L} y(i) \Big/ L$$

$x(i)$: pitch peak of word *i* uttered by a native English speaker
$y(i)$: pitch peak of word *i* uttered by a native Japanese speaker
L: number of word in a sentence

The mean values of individual words for both groups of the Japanese English and the Native English are obtained. We denote these mean values by *Ave*, which are calculated as follows:

$$\begin{aligned} x(i)_{ave} &= \sum_{i=1}^{N} x(i)' \Big/ N \\ y(i)_{ave} &= \sum_{i=1}^{M} y(i)' \Big/ M \end{aligned} \tag{2}$$

N: number of males or females of English speakers
M: number of males or females of Japanese speakers
(Henceforth, the $x(i)_{ave}$ and $y(i)_{ave}$ are considered as equivalent to *Ave*.)

The large value of *Ave* indicate high pitch peak. In this study, "Ave>1.1" indicates that *Ave* is higher than 1.1, and "Ave<0.9" indicates that *Ave* is lower than 0.9. For each group, the *Ave* of words for the cases in which "Ave>1.1" and "Ave<0.9" are counted.

2.2.2. R value

Statistical significance of the difference between $x(i)'$ and $y(i)'$ can be evaluated by criterion used in statistical pattern recognition, that is, a ratio of between-group variance to within-group variance. We denote this ratio by R, as in equation (3). If R is large, it indicates that considerable difference exists in sample

distributions of the two groups.

$$R = \frac{(\overline{x} - \overline{y})}{(\sigma_x + \sigma_y)} \quad (3)$$

where

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x(i)', \qquad \overline{y} = \frac{1}{M}\sum_{i=1}^{M} y(i)'$$

$$\sigma_x = \frac{1}{N}\sum_{i=1}^{N}(x(i)' - \overline{x})^2, \quad \sigma_y = \frac{1}{M}\sum_{i=1}^{M}(y(i)' - \overline{y})^2$$

In this study, "ntv>jpe" indicates that $peak(i)'$ of the corresponding word for a group of Native English is higher than these for a group of Japanese English, and "ntv<jpe" indicates the reverse. Here, difference of distinction between male and female is considered, and only the case of significant difference between the two groups is counted. Specifically, inequality, "ntv>jpe" or "ntv<jpe", holds only for cases that satisfy $R > 0.2$ for male and $R > 0.3$ for female are counted.

2.2.3. Max, Min and Dynamic range

The word whose pitch peak is the highest in a sentence is referred to as *Max*. The word whose pitch peak is the lowest in a sentence is referred to as *Min*. In this study, if a *Max* or a *Min* for Japanese English is equal to either male or female of the corresponding word for Native English, it is determined to be "same". "differ" indicates that a *Max* or a *Min* for Japanese English differs from neither male nor female of the corresponding word for Native English. Then the cases that satisfy "same" or "differ" are counted.

Then dynamic range of pitch peaks of words in a sentence (*D-range*, henceforth) is calculated as follows:
$D - range = Max - Min$

The larger the value of *D-range* is, the larger is the intonation of the sentence. In this study, "ntv>jpe" indicates that *D-range* for Native English is larger than that for Japanese English, and "ntv<jpe" indicates the reverse. The cases that satisfy "ntv>jpe" or "ntv<jpe" are counted.

3. Speech Used in Analysis

3.1. Subjects

3.1.1. Native English speakers

The group of native English speakers consisted of 10 subjects (five male, five female), aged between 20 and 40. Most were English teachers living in Japan, from the United Kingdom (4), Canada (2), New Zealand (2), Australia (1), and the United States (1).

3.1.2. Native Japanese speakers

This group consisted of 17 subjects (nine male, eight female), aged between 20 and 30. Most were undergraduate students.

A native speaker of English, who was an English teacher in Japan, listened to the utterances of all subjects, and judged their levels of English prosody. The prosodies in their English utterances are judged as the typical ones for the majority of Japanese learners of English. (i.e. They are not very good at English.)

3.2. Sample Sentence

One hundred sentences (84 declarative, 16 interrogative) were chosen from the MOCHA-TIMIT data set (timit001-030, 211-260, 441-460), which exhibits considerable variation in meaning and structure, including five passive sentences, four comparative sentences and five negative sentences. Each sentence was composed of three to 12 words and there were 707 words in total.

3.3. Recording Condition

The subjects were given sufficient time to practice reading the speech materials before recording. They were also 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 Native English each uttered 100 sentences. One group of nine (five male, four female) Japanese English each uttered 50 sentences (timit001-030, 211-230). A second group of eight Japanese English each uttered the remaining 50 sentences.

4. Results

1414 Ave values, from Ave_1 to Ave_{707}, and 1414 R values, from R_1 to R_{707}, are obtained from 707 words in total males and females analyses. Words are divided into word class and into position within a sentence.

Individual words are classified into content words and function words, where there exist 451 content words and 256 function words. Content words are further classified into noun, verb, adjective and adverb. Function words are also further classified as conjunction/preposition, be/auxiliary verb/do, article, pronoun and interrogative/negative.

Words are divided into following three parts: 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 does not end the sentences, are included in words at the ends of sentences. This study defines these as second ends.

Results are summarized as follows:

4.1. Word Position

4.1.1. Beginning

There are 100 words at the beginning of sentences, which include 38 content words, represented by 'C', and 62 function words, 'F'. Table 1 shows the results of *Ave* for words at the beginning of sentences. From the table, we can see that content words of males and females amount to 76. For 76 content words for Native English, 36% of which satisfy "$Ave > 1.1$". For Japanese English, only 17% of which satisfy "$Ave > 1.1$".

Table 2 shows the results of *R* for words at the beginning of and within sentences, represented by 'beginning' and 'within', respectively. Out of 76 content words at the beginning of sentences, 40 of the words satisfy "$R > 0.2$ or 0.3", 78% of which satisfy "$ntv > jpe$".

These mean that the pitch peaks of content words at the beginning of sentences for Japanese English are lower than those for Native English.

For 124 function words for Native English in Table 1, 51% of which satisfy "$Ave < 0.9$". For Japanese English, only 24% of which satisfy "$Ave < 0.9$". In addition, out of 124 function words in Table 2, 62 of the words satisfy "$R > 0.2$ or 0.3", 77%

of which satisfy "$ntv < jpe$". These mean that the pitch peaks of function words at the beginning of sentences for Japanese English are higher than those for Native English.

Table 1: *Results of* Ave *of words at the beginning of sentences.*

	ntv		jpe	
	C	F	C	F
number of word	76	124	76	124
$Ave > 1.1$	27 (36%)	7	13 (17%)	1
$Ave < 0.9$	4	64 (51%)	4	30 (24%)
$0.9 \leqq Ave \leqq 1.1$	45	53	59	93

Table 2: *Results of* R *of words at the beginning of and within sentences.*

	beginning		within	
	C	F	C	F
number of word	76	124	616	378
$R > 0.2$ or 0.3	40	62	315	227
$ntv > jpe$	31 (78%)	14	183 (58%)	31
$ntv < jpe$	9	48 (77%)	132	196 (86%)
$R \geqq 0.2$ or 0.3	36	62	301	151

4.1.2. Within

There are 994 words within sentences, which include 616 content words and 378 function words. The results of *Ave* and *R* for words within sentences are shown in Table 3 and Table 2, respectively. For content words within sentences, results suggest that the pitch peaks for Japanese English are a little lower than those for Native English. For function words within sentences, results suggest that the pitch peaks for Japanese English are higher than those for Native English.

Table 3: *Results of* Ave *of words within sentences.*

	ntv		jpe	
	C	F	C	F
number of word	616	378	616	378
$Ave > 1.1$	182 (30%)	4	110 (18%)	
$Ave < 0.9$	51	196 (52%)	24	78 (21%)
$0.9 \leqq Ave \leqq 1.1$	383	178	482	300

4.1.3. End

There are 200 words at the ends of sentences. The words at the ends of sentences are classified as words appearing in declarative sentences or appearing in interrogative sentences: there are 168 words at the ends of declarative sentences 'dec' and 32 words at the ends of interrogative sentences 'int'. There are 20 second ends 'sec'. The results of *Ave* and *R* for words at the ends of sentences are shown in Table 4 and Table 5, respectively.

For words at the ends of declarative sentences, the results suggest that the pitch peaks of the words for Japanese English are higher than those for Native English. However, for words at the ends of interrogative sentences and second end, the results suggest that the pitch peaks of the words for Japanese English are lower than those for Native English.

Figure 1 shows mean values and standard deviations of $peak(i)'$ for a sentence example. The sentence is timit248, "*They all agree that the essay is barely intelligible*". The function word at the beginning of the sentence *They*$_{(R=0.97)}$ satisfies "ntv<jpe", where R value is given in parenthesis. The content words within the sentence *all*$_{(2.27)}$ and *agree*$_{(5.96)}$ satisfy "ntv>jpe". The function words within the sentence *that*$_{(1.97)}$ and *is*$_{(0.33)}$ satisfy "ntv<jpe". The word at the end of the declarative sentence *intelligible*$_{(1.89)}$ satisfy "ntv<jpe".

Table 4: *Results of* Ave *of words at the ends of sentences.*

	ntv			jpe		
	int	dec	sec	int	dec	sec
number of word	32	168	20	32	168	20
$Ave > 1.1$	19 (59%)	36 (21%)	13 (65%)	13 (41%)	61 (36%)	7 (35%)
$Ave < 0.9$		3				
$0.9 \leqq Ave \leqq 1.1$	13	129	7	19	107	13

Table 5: *Results of* R *of words at the ends of sentences.*

	int	dec	sec
number of word	32	168	20
$R > 0.2$ or 0.3	16	73	12
$ntv > jpe$	13 (81%)	18	10 (83%)
$ntv < jpe$	3	55 (75%)	2
$R \geqq 0.2$ or 0.3	16	95	8

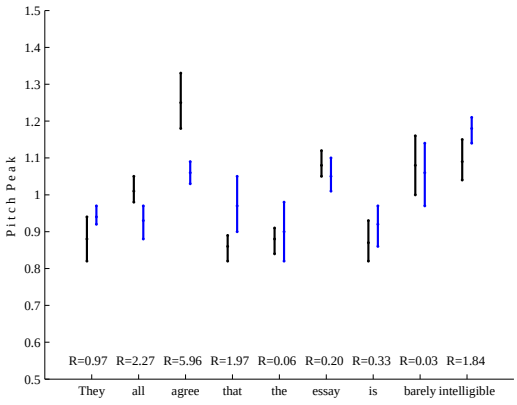

Figure 1: $peak(i)'$ *distribution of a sentence example for English speakers (right-side bar) and Japanese speakers (left-side bar). Each bar indicates the range from (mean - SD) to (mean + SD)*

4.2. Word Class

4.2.1. Content Word

Content words collectively in male and female speeches are 378 noun 'noun', 270 adjective 'adj', 212 verb 'verb' and 42 adverb 'adv'. Table 6 shows the results of R for content words. The results suggest that for the nouns and adjectives, more than half of which satisfy "ntv>jpe". For the adverbs and verbs, more than half of which satisfy "ntv<jpe".

Fig. 2 shows an analysis example for timit003, *"Those thieves stole thirty jewels"*. *thieves*$_{(R=1.4)(noun)}$ satisfies "ntv>jpe", where word class is given in the second parenthesis. On the contrary, *stole*$_{(0.86)(verb)}$ satisfies "ntv<jpe".

Table 6: *Results of* R *of content words.*

	noun	adj	verb	adv
number of word	378	270	212	42
$R > 0.2$ or 0.3	188	136	111	17
$ntv > jpe$	121 (64%)	74 (54%)	51	8
$ntv < jpe$	67	62	60 (54%)	9 (52%)
$R \geqq 0.2$ or 0.3	190	134	101	25

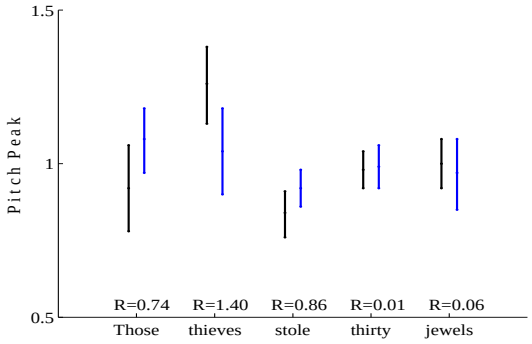

Figure 2: $peak(i)'$ *distribution of content words for English speakers and Japanese speakers.*

4.2.2. Function Word

Function words are also further classified as 168 conjunction/preposition 'cnj', 74 be/auxiliary verb/do 'be', 148 article 'art', 98 pronoun 'prn' and 24 interrogative/negative 'inng'. Table 7 shows the results for R values of function words. For conjunctions/prepositions, be/auxiliary verb/do, articles and pronouns, results suggest that the pitch peaks of those words for Japanese English are higher than those for Native English. On the contrary, results suggest that the pitch peaks of interrogatives/negatives for Japanese English are lower than those for Native English.

Figure 3 shows an analysis example for timit251, "*How ancient is this subway escalator*". *How*$_{(0.32)(interrogative)}$ satisfies "ntv>jpe", however *is*$_{(1.22)(be)}$ satisfies "ntv<jpe".

4.3. Max, Min and D-range

200 Max values, from Max_1 to Max_{100}, 200 Min values, from Min_1 to Min_{100} and 200 $D-range$ values, from $D-range_1$ to $D-range_{100}$ are obtained from 100 sentences in total.

Table 7: *Results of* R *of function words.*

	inng	cnj	be	art	prn
number of word	24	168	74	148	98
$R > 0.2$ or 0.3	15	97	37	85	59
$ntv > jpe$	14 (93%)	11	6	6	9
$ntv < jpe$	1	86 (89%)	31 (84%)	79 (93%)	50 (85%)
$R \geqq 0.2$ or 0.3	9	71	37	63	39

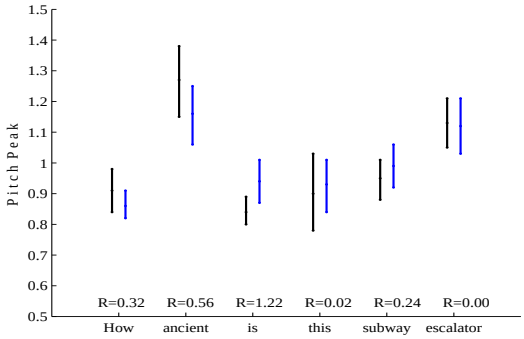

Figure 3: $peak(i)'$ *distribution of function words for English speakers and Japanese speakers.*

For 200 *Max*, 115 (58%) of which satisfy "same". For 200 *Min*, 110 (55%) of which satisfy "differ".

For 200 *D-range*, 171 (88%) of which satisfy "ntv>jpe". Therefore, *D-range* for Japanese English is narrower than that for Native English.

5. Conclusion

The pitch peaks of words in sentences were compared for Native English and Japanese English. First, results show that the content words at the beginning of sentences for Japanese English have lower pitch than those for Native English. On the contrary, the end words of declarative sentences have higher pitch than those for Native English. Additionally, pitch ranges of sentences in Japanese English are narrower than those for Native English. Taken together, these suggest that intonation patterns in sentence utterances by Japanese English are rather flat.

Furthermore, results suggest that the majority of function words for Japanese English have higher pitch than those for Native English. Additionally, the end words at interrogative sentences and subordinate sentences in Japanese English have lower pitch than for Native English.

6. References

[1] M. Kondo, "Is acquisition of L2 phonemes difficult? Production of English stress by Japanese speakers," In Proc. the 10th Generative Approaches to Second Language Acquisition Conference, pp105-112, 2009.

[2] Y. Takehuta, Nihonjin Eigo no Kagaku, Kenkyusha shuppan, Tokyo, 1982. (in Japanese)

[3] H. Obari, R. Tomiyama, M. Yamamoto and S. Itahashi, "Differentiation of English utterances of Japanese and native speakers by several prosodic parameters," Proc. Oriental COCOSDA, pp.143-147. 2005.