PAPER

A Study on Pitch Patterns in Japanese Speakers of English with Verification by Speech Re-Synthesis

Tomoko NARIAI[†a)], *Nonmember and* **Kazuyo TANAKA**[†], *Fellow*

SUMMARY Certain irregularities in the utterances of words or phrases often occur in English spoken by Japanese native subject, referred to in this article as Japanese English. Japanese English is linguistically presumed to reflect the phonetic characteristics of Japanese. We consider the prosodic feature patterns as one of the most common causes of irregularities in Japanese English, and that Japanese English would have better prosodic patterns if its particular characteristics were modified. This study investigates prosodic differences between Japanese English and English speakers' English, and shows the quantitative results of a statistical analysis of pitch. The analysis leads to rules that show how to modify Japanese English to have pitch patterns closer to those of English speakers. On the basis of these rules, the pitch patterns of test speech samples of Japanese English are modified, and then re-synthesized. The modified speech is evaluated in a listening experiment by native English subjects. The result of the experiment shows that on average, over three-fold of the English subjects support the proposed modification against original speech. Therefore, the results of the experiments indicate practical verification of validity of the rules. Additionally, the results suggest that irregularities of prominence lie in Japanese English sentences. This can be explained by the prosodic transfer of first language prosodic characteristics on second language prosodic patterns.

key words: *second language learning, Japanese learners of English, prominence, analysis by synthesis, prosody*

1. Introduction

English is studied as a second language in junior high school or high school in Japan. However, we, Japanese, often have difficulty in making ourselves understood in English when we actually talk to a native English speaker. There are certain differences in utterances between English speakers and Japanese speakers. Thus, we have the phenomenon known as *Japanese English*.

This has been a matter of some concern to people involved in studying ways to improve Japanese English. The number of computer-based studies of Japanese English has increased markedly over the last decade, where analyses of the production of English phonemes have been conducted [1]. The obtained knowledge deploys speech technology into computer-based educational systems that can be used to teach foreign language skills [2], [3].

Although a large body of research on second language production has been conducted in the phonemic domain, correct usage of prosodic patterns has been shown to improve the syntactic and semantic intelligibility of spoken language [4]. Therefore, there has been a strong research interest in identifying the prosodic characteristics of Japanese English [5], [6]. However, few studies on Japanese English have examined the production of the prosodic patterns in the acoustic-phonetic level. In addition, most past studies have following problems.

First, few studies come up with concrete proposals for improving the characteristics in Japanese English; none of these characteristics have yet been confirmed in terms of actual speech modification. For example, previous study [7] analyzed the range of maximal and minimal pitch in Japanese English sentences. It was revealed that the dynamic range of pitch in Japanese English sentences was smaller than in English speakers sentences. On the basis of this finding, we tried to improve the characteristics of Japanese English by means of a speech synthesizer. This type of modification, however, cannot cover the gap in pitch between Japanese English sentences and English speakers sentences.

It suggests that the statistical difference found by a bottom-up approach will not suggest how Japanese English should be modified to have pitch patterns closer to those of English speakers. As a method of research, knowledge extraction by analysis with verification by re-synthesis is both informative and reliable in investigating second language prosody.

Second, our previous study revealed that defects of focus or prominence occurred in lengthening in Japanese English [8]. Prominence in English emerges primarily as a change in pitch (i.e., acoustically in fundamental frequency), and secondarily as a longer duration of a word [9]. Therefore, irregularities of prominence in Japanese English are presumed to occur in the pitch pattern associated with prominence.

Finally, the result of a previous study [7] suggests that focusing on only a whole sentence in Japanese English sentences may lead to oversimplification. A detailed investigation should be made of what sentence units should be crucial. In this study, Japanese English sentences are investigated by using sentence structure to classify words depending on their roles or functions.

In this study, the prosodic features in Japanese English are analyzed on the basis of a prediction from two language systems, and examined by speech re-synthesis.

The first half of the paper describes the pitch patterns of Japanese English sentences by comparative analysis between English speakers and Japanese English. We ana-

Manuscript received April 14, 2011.
Manuscript revised July 19, 2011.

[†]The authors are with the Graduate School of Library, Information and Media Studies, University of Tsukuba, Tsukuba-shi, 305–0085 Japan.
a) E-mail: nariai@slis.tsukuba.ac.jp

DOI: 10.1587/transinf.E94.D.2495

lyze pitch of Japanese English on the assumption that the prosodic difference between the English and Japanese languages appears in prominence. In the latter half, modification rules are derived from the analytical results, which prove the peculiarities of the pitch patterns of Japanese English. The rules are acoustically realized by speech resynthesis, and then evaluated by a listening experiments.

Section 2 describes speech samples and analysis method. Sentences of speech samples are divided into words for the analysis with regard to the word class. In Sect. 3, the analysis of words in Japanese English sentences is conducted in comparison with in English speakers sentences. In Sect. 4, the way of modifying pitch patterns of Japanese English is described as rules. On the basis of the rules, Japanese English is acoustically modified. The modified speech is evaluated in a listening experiment taken by English speakers. A discussion of the analytical results is presented in Sect. 5, and the conclusion is given in Sect. 6.

2. Speech Samples and Analysis Method

2.1 Sample Sentences

The sentence text set of this speech dataset is the same as that of MOCHA-TIMIT dataset [10]. In the analyses, 100 sentences are chosen; the sentence numbers are timit001-030, 211-260 and 441-460. There are 707 words in total.

2.2 Subjects

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

The group of Japanese English consisted of 17 subjects, 9 males and 8 females, aged between 20 and 30. Most of the subjects were undergraduate students. A native speaker of English, who is an English teacher in Japan, listened to the utterances of all subjects. They were judged not to be so 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 Outline of Analysis Method

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 pitch patterns of individual sentences are segmented into word sequences, where word boundaries are determined by observing the waveform and the spectrogram pattern. Words with articulation errors are not deleted as long as they do not interfere in word boundary detection.

The values of $peak(i)$ and $range(i)$ of individual words i are estimated as characterizing its prosodic patterns. The values are defined as:

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

$range(i)$ of pitch =*maximal minus minimal fundamental frequency of word i*

In this paper, the fundamental frequency is treated in a linear scale domain.

2.5 Statistical Measure Used in the Analysis

Statistical significance of the difference in sample distributions between the two groups can be evaluated by criterion used in statistical pattern recognition, that is, a ratio of the between-group variance to the within-group variance, known as Fisher's ratio in linear discriminant analysis. This ratio is denoted by R. If R is large, it indicates that considerable difference exists in sample distributions of the two groups. A procedure to calculate R is as follows:

(1) Prosodic features of each word are normalized by the average of those values of the words contained in the corresponding utterance of a sentence. That is, average feature values of word i for each sentence are represented by the following equations:

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

$$\overline{y}_j = \sum_{i=1}^{L} y_j(i) \Big/ L \tag{2}$$

$x_j(i)$: prosodic feature value of word i uttered by an English speaker j

$y_j(i)$: prosodic feature value of word i uttered by a Japanese speaker j

L: number of words contained in the corresponding sentence

Then, relative feature values of word i are written as follows:

$$x_j(i)' = x_j(i) \Big/ \overline{x}_j \tag{3}$$

$$y_j(i)' = y_j(i) \Big/ \overline{y}_j \tag{4}$$

(2) Calculate mean values and variances of $x_j(i)'$ and $y_j(i)'$ for English subjects and Japanese subjects, as follows:

$$\overline{x}(i) = \frac{1}{N} \sum_{j=1}^{N} x_j(i)' \tag{5}$$

$$\overline{y}(i) = \frac{1}{M}\sum_{j=1}^{M} y_j(i)' \tag{6}$$

$$\sigma_x(i) = \frac{1}{N}\sum_{j=1}^{N}(x_j(i)' - \overline{x}(i))^2 \tag{7}$$

$$\sigma_y(i) = \frac{1}{M}\sum_{j=1}^{M}(y_j(i)' - \overline{y}(i))^2 \tag{8}$$

N: number of English speakers
M: number of Japanese speakers
(3) Thus R of word i is obtained as follows:

$$R(i) = \frac{(\overline{x}(i) - \overline{y}(i))^2}{\left(\sigma_x(i) + \sigma_y(i)\right)} \tag{9}$$

R values are calculated for individual words. Parameters $peak(i)$ and $range(i)$ are redefined using normalized prosodic patterns, $x_j(i)'$ and $y_j(i)'$, as

$peak(i)'$: maximal value of $x_j(i)'$ (or $y_j(i)'$)
$range(i)'$: maximum minus minimum of $x_j(i)'$ (or $y_j(i)'$)

In the analysis in Sect. 3, R>0.1 is used for an index to detect the difference between Japanese subjects and English subjects. In the analysis, "ntv>jpe" indicates that $peak(i)'$ or $range(i)'$ of pitch for English subjects "ntv" is higher or larger than that for Japanese subjects "jpe," and "ntv<jpe" indicates the reverse.

2.6 Word Class

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, including the present progressive forms and passive verbs), adjectives (play the role of adjective in sentences), and adverbs (play the role of adverb in sentences).

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

3. Analysis of Pitch Patterns of Japanese English

3.1 Pitch Peak

This subsection investigates $peak(i)'$ of pitch depending on the word class. Number of words both of male and female amounts to 1414 in total, which are divided into content words and function words.

Table 1 shows the results of $peak(i)'$ for content words. From the table, we can see that noun of males and females amounts to 380, represented by 'noun', adjectives to 270, 'adj', verbs to 210, 'verb', and adverbs to 44, 'adv'. For each word class, the words satisfying "ntv>jpe" and "ntv<jpe" are counted.

Out of 380 nouns, 244 satisfy $R > 0.1$, 61% of which satisfy "ntv>jpe". Out of 270 adjectives, 160 satisfy $R > 0.1$, 53% of which satisfy "ntv>jpe". In contrast, out of 210 verbs, 144 satisfy $R > 0.1$, 53% of which satisfy "ntv<jpe". Out of 44 adverbs, 26 satisfy $R > 0.1$, 58% of which satisfy "ntv<jpe".

Table 1 Results of $peak(i)$ of pitch for content words.

	noun	adj	verb	adv
number of words	380	270	210	44
$R > 0.1$	244	160	144	26
ntv>jpe	150 (61%)	84 (53%)	67	11
ntv<jpe	94	77	77 (53%)	15 (58%)

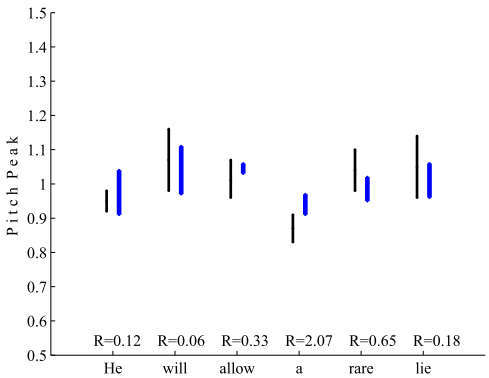

Fig. 1 Pitch peak of $peak(i)'$ distribution of words in timit011 for English subjects (left-side thin bar) and Japanese subjects (right-side bold bar). Each bar indicates the range from (*mean* - *SD*) to (*mean* + *SD*).

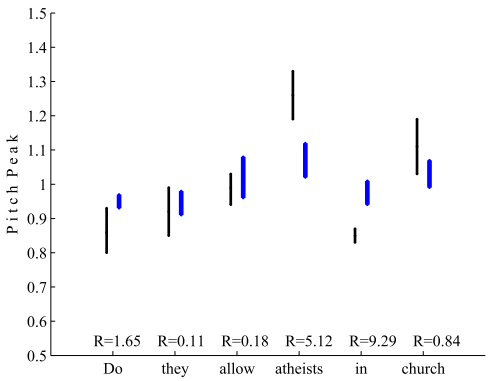

Fig. 2 $peak(i)'$ of pitch in timit259 for English and Japanese speakers.

In Fig. 1, the mean and standard deviation of $peak(i)'$ of words in timit011, "*He will allow a rare lie*", are plotted for English and Japanese speakers. The words that satisfy "ntv>jpe" are $rare_{(R=0.65)(adjective)}$ and $lie_{(0.18)(adjective)}$, where the R value for each word is given in first parenthesis and the word class is given in the second parenthesis. The word for "ntv<jpe" is $allow_{(0.33)(verb)}$.

Figure 2 shows the result for timit259, "*Do they allow atheists in church*", where $atheists_{(5.12)(noun)}$ and $church_{(0.84)(noun)}$ satisfy "ntv>jpe", however, $allow_{(0.18)(verb)}$

Table 2 Results of *peak*(*i*) of pitch for function words.

	int, ng	cnj, prp	be	art	prn
number of words	22	168	74	148	98
$R > 0.1$	15	118	49	108	71
ntv>jpe	14 (93%)	15	7	11	12
ntv<jpe	1	103 (87%)	42 (86%)	97 (90%)	59 (83%)

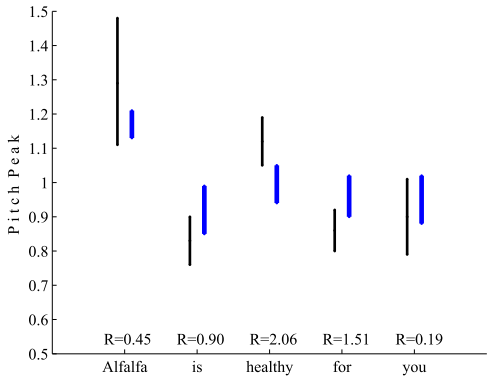

Fig. 3 *peak*(*i*)′ of pitch in timit021 for English and Japanese speakers.

satisfies "ntv<jpe".

Table 2 shows the results for function words. Function words of males and females amount to 512 words: 22 interrogatives/negatives, 'int, ng'; 168 conjunctions/prepositions, 'cnj, prp'; 74 be/auxiliary verb/do, 'be'; 148 articles, 'art'; and 98 pronouns, 'prn'.

Out of 22 interrogatives/negatives, 15 satisfy $R > 0.1$. Of these 15, 93% of which satisfy "ntv>jpe". On the contrary, over 80% of the conjunctions/prepositions, be/auxiliary verb/do, article, and pronoun satisfy "ntv<jpe".

These suggest that most function words for Japanese English have higher pitch than those for English speakers. Also, interrogative/negative for Japanese English has lower pitch than that for English speakers.

Figure 3 shows the result of words in timit021, "*Alfalfa is healthy for you*," where $is_{(0.90)(be)}$, $for_{(1.51)(preposition)}$ and $you_{(0.19)(pronoun)}$ satisfy "ntv<jpe."

Figure 4 shows the result of words in timit229, "*Keep the thermometer under your tongue*," where $the_{(0.37)(article)}$, $under_{(0.26)(preposition)}$ and $your_{(0.30)(pronoun)}$ satisfy "ntv<jpe."

3.2 Pitch Range

This subsection investigates the *range*(*i*)′ of pitch depending on the word class.

Table 3 shows the results of *range*(*i*)′ of pitch for content words. The results suggest that more than half of nouns and adjectives satisfy "ntv>jpe." In contrast, for verbs and adverbs, more than half of which satisfy "ntv<jpe."

These suggest that nouns and adjectives for Japanese English have smaller pitch range than those for English speakers.

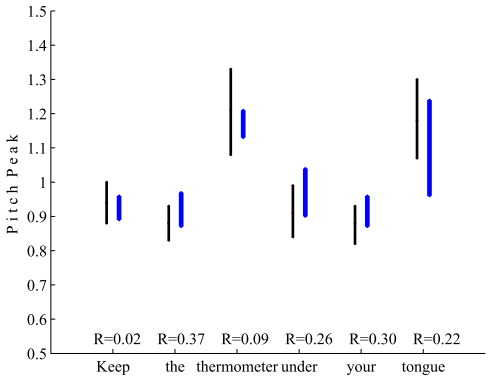

Fig. 4 *peak*(*i*)′ of pitch in timit229 for English and Japanese speakers.

Table 3 Results of *range*(*i*) of pitch for content words.

	noun	adj	verb	adv
number of words	380	270	210	44
$R > 0.1$	237	193	138	31
ntv>jpe	129 (54%)	107 (55%)	61	15
ntv<jpe	108	86	77 (56%)	16 (52%)

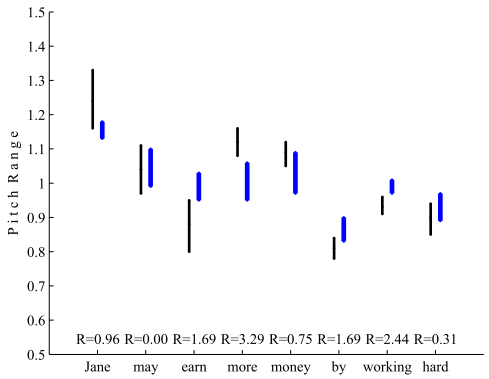

Fig. 5 *range*(*i*)′ of pitch in timit004 for English and Japanese speakers.

Figure 5 shows the result of words in timit004, "*Jane may earn more money by working hard*," where $Jane_{(0.96)(noun)}$, $more_{(3.29)(adjective)}$ and $money_{(0.75)(noun)}$ satisfy "ntv>jpe," however, $earn_{(1.69)(verb)}$, $working_{(2.44)(verb)}$ and $hard_{(0.31)(verb)}$ satisfy "ntv<jpe."

Figure 6 shows the result of words in timit241, "*Clear pronunciation is appreciated*," where $Clear_{(1.10)(adjective)}$ and $pronunciation_{(0.36)(noun)}$ satisfy "ntv>jpe," however, $appreciated_{(0.56)(verb)}$ satisfies "ntv<jpe."

Table 4 shows the results of *range*(*i*)′ for function words. The results suggest that the majority of function words satisfy "ntv<jpe." However, only for interrogatives/negatives, the majority satisfies "ntv>jpe."

These suggest that most function words for Japanese English are larger pitch range than those for English speakers. Also, interrogative/negative for Japanese English is smaller pitch range that than that for English speakers.

Figure 7 shows the result of words in timit221, "*How*

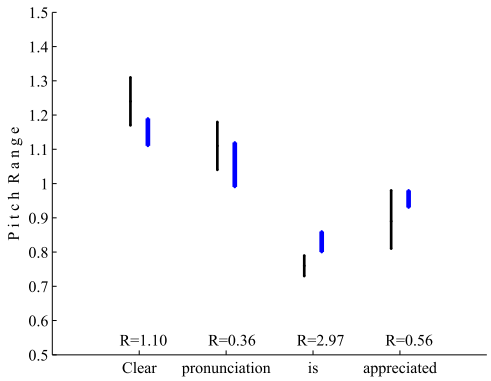

Fig. 6 *range*(i)′ of pitch in timit241 for English and Japanese speakers.

Table 4 Results of *range*(i)′ for Function Words.

	int, ng	cnj, prp	be	art	prn
number of words	22	168	74	148	98
$R > 0.1$	18	125	53	96	56
ntv>jpe	18 (100%)	17	22	35	14
ntv<jpe		108 (86%)	31 (58%)	61 (64%)	42 (75%)

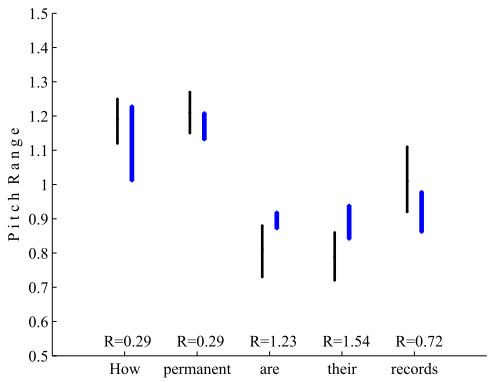

Fig. 7 *range*(i)′ of pitch in timit221 for English and Japanese speakers.

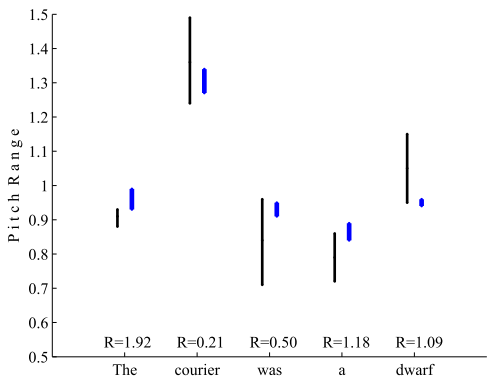

Fig. 8 *range*(i)′ of pitch in timit242 for English and Japanese speakers.

permanent are their records," where $How_{(0.29)(interrogative)}$ satisfies "ntv>jpe," however, $are_{(1.23)(be)}$ and $their_{(1.54)(pronoun)}$ satisfy "ntv<jpe."

Figure 8 shows the result of words in timit242, "*The courier was a dwarf*," where $The_{(1.92)(article)}$, $was_{(0.50)(be)}$ and $a_{(1.18)(article)}$ satisfy "ntv<jpe."

4. Verification of Sentence Pitch Pattern Features in Japanese English by Speech Re-synthesis

The analytical results in the previous section point to new ways, in which the irregularities of the pitch patterns in Japanese English are improved. In this section, the rules are derived based on the analytical results. Then, the modification method of Japanese English samples and the analytical result are described.

4.1 Method

We create two rules to improve irregularities of pitch patterns in Japanese English: One is for each word; the Second is for each sentence.

For the former, a rule is derived by the results of pitch peak and pitch range for words in Japanese English sentences with regard to the word class.

For the latter, it is known that when English sentences are spoken aloud, they are broken into smaller phrases, that correspond to lexical or phonetic units [11], [12]. Each end of phrase is indicated by a decline in pitch, as is stated in many as a phenomenon in English utterance. It is, however, reported that less proficient second language speakers tend to divide utterances into smaller phrases [5], so it is expected that phrase size will influence prosodic grouping of speech in second language speech. Therefore, the irregularities of phrasing in Japanese English are considered to affect the result of pitch at end words in Japanese English sentences. Therefore, phrasing is checked whether at the ends of declarative sentences are lower.

The check points are whether pitch patterns for words in Japanese English sentences are arranged properly, and whether an accent phrasing in Japanese English forms properly at the end of the sentence. Two rules to modifying way of pitch patterns of Japanese English can be stated as follows.

Japanese English will have improved pitch patterns if:

1. *peak*(i) and *range*(i) of each word in a sentence are ordered as follows:
   [*rule1*] *function word* < (*verb, adverb*) < (*noun, adjective, interrogative, negative*)
2. *peak*(i) of the word at the end of sentence has the lowest pitch, as follows:
   [*rule2*] *the end word* < *the word within a sentence*

Japanese English samples are modified to adjust to the rules, if there includes an erroneous order. Japanese English samples are analyzed to list what needs to be modified. The analysis process has the following five steps:

Table 5 List of irregularities of Japanese English.

sample number	detected rule number
sample-1	rule-1
sample-2	rule-2
sample-3	rule-1
sample-4	rule-1
sample-5	rule-1 and -2
sample-6	rule-1 and -2

(i) A speech sample of Japanese English is analyzed by STRAIGHT [13] to extract the pitch patterns. The pitch patterns are manually aligned with the word boundaries.
(ii) *peak*(i) and *range*(i) for words in a sentence are measured.
(iii) *peak*(i) and *range*(i) of words in the sentence are ranked according to its values. The ranking is compared with *rule 1*. Then, the irregularities are detected.
(v) The pitch height is checked by *rule 2*, and then, the irregularities are detected.

4.2 Sample Speech

Six Japanese subjects (four males, two females), aged between 20 and 30, were chosen. Most were Japanese university students. A native speaker of English judged that they were not so proficient in English.

Six sample sentences were chosen at random from the MOCHA-TIMIT data set, the sentence numbers of which are timit 009, 021, 022, 216, 246 and 452.

Each subject was allocated a different sentence, which they uttered once. Subjects were assigned number sample-1 to sample-6.

4.3 Analysis Result

Speech samples are analyzed to list which rules need to be fixed. The list of irregularities of sample-1 to sample-6 is indicated in Table 5.

4.4 Modification Method

The pitch patterns are modified as accurate as possible to adapt the rules given in subsection 4.1. Speech signals with those pitch patterns are re-synthesized by STRAIGHT.

The pitch patterns of a word were modified according to the following equation.

$$\widetilde{f_0}(t) = f_{mean} + (f_0(t) - f_{mean}) \times a + b \qquad (10)$$

$f_0(t)$: *pitch frequency pattern*
a: *dynamic range modification factor*
b: *peak adjustment factor*

where f_{mean} denotes the mean value of the pitch patterns of a corresponding word. a is a parameter for amplification of the selected pitch pattern. If $a > 1$, then the pitch range is amplified. b is a parameter that boost (plus) or depress (minus) the selected pitch pattern. If a word does not need modification, $(a, b) = (1, 0)$ is used.

Table 6 Modification parameters a and b used for the experiment.

defect	word class	a, b
pitch range and pitch peak	function word	$0 < a < 1, -30 < b < 20$
	adverb,verb	$0.5 < a < 1.5, -10 < b < 30$
	noun,adjective	$1 \le a < 1.5, -50 < b < 40$
pitch fall	word within sentence	$0.5 < a < 1.5, -10 < b < 30$
	end word	$1 \le a < 1.5, -50 < b < 40$

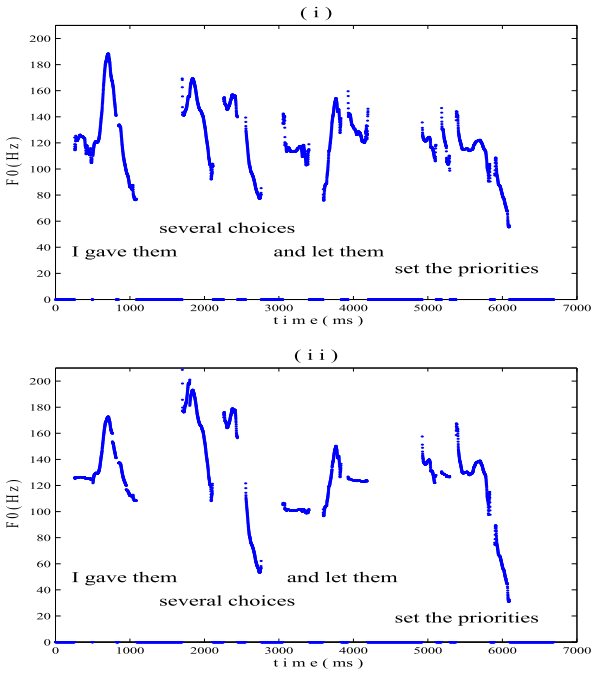

Fig. 9 Pitch patterns of sample-4 of (i) the original speech and (ii) modified speech.

The irregularities of pitch patterns are improved by a and b as in Table 6. For modification of *rule 1* and *rule 2*, a is mainly used, and additionally b is used to produce a proper balance.

The irregularities are improved in pitch range (*rule1*) and pitch fall (*rule2*) by a and b shown in Table 6.

Figure 9 shows the contrasting pitch patterns of *sample-4*: "*I gave them several choices and let them set the priorities*," where (i) and (ii) illustrate the pitch pattern of the original and modified speech, respectively.

The pitch patterns of all six speech samples listed in Table 5 are modified in the same manner as the above.

4.5 Listening Experiment

4.5.1 Subjects

The subjects for evaluating speech samples were 18 native English speakers (3 males, 15 females), aged between 19 and 40. Most were undergraduate or graduate students in Michigan.

Table 7 Results of listening experiment.
S:support, N: not support I: Indefinite

	S	N	(I)
sample-1	**10**	2	(6)
sample-2	**11**	2	(5)
sample-3	**11**	5	(2)
sample-4	**10**	4	(4)
sample-5	**14**	2	(2)
sample-6	**9**	6	(3)

4.5.2 Procedure

The modified speech is examined using an evaluation test, in which a pair of contrasting speech samples, the original one and its modification, i.e., those shown in Fig. 9 (i) and (ii), are presented randomly to subjects.

The test was carried out in a quiet room. The subjects were requested to listen to a pair of original and modified speech samples, then to answer the following question: "*Which sample of the two had more natural pitch patterns in English.*" The subjects were instructed to answer *I* if he or she could not catch the difference in pitch patterns of the two contrastive speech samples, or could not decide which should be chosen.

4.5.3 Result

The results of the listening experiment are shown in Table 7, where *S* indicates an answer that supports the modified speech, *N* indicates one that does not support the modified speech, and *I* indicates that the subject could not distinguish between the contrasting speech samples.

Table 7 shows that the modified speech sounds more natural to the native English speakers than the original versions. This is true for all six samples, and averagely, over threefold subjects support the proposed modification against the original speech. Therefore, our approach is considered to be practically verified.

5. Discussion

Our previous study [8] for the durations of Japanese English and those of English speakers indicated that the duration of nouns for Japanese English was much shorter than for English speakers. For the analysis of pitch, however, there was little difference among content words. This shows that irregularities of prominence in Japanese English occur frequently in the lengthening, rather than pitch. In this study, however, the pitch patterns in Japanese English were analyzed, and then examined by speech re-synthesis. Therefore, we can confirm irregularities of pitch prominence in Japanese English.

The result of this study supported the following three generalizations:

First, it confirms the knowledge indicated in previous studies that Japanese English tends not to emphasize important words (i.e., content word) [6]. In addition, our results add the knowledge that important words are nouns, adjectives, interrogative and negative.

Second, the result contributes to the advancement in constructing extensible language learning systems [2], [14], by establishing that English speakers put prominence on the stressed syllable in important words, whereas Japanese do not.

Finally, the irregularities indicated in this study can be explained by the prosodic transfer of first language prosodic characteristics on second language prosodic patterns [15], [16]. In English, prominence is assigned to words that bear the primary meaning [17], [18]. On the other hand, prominence in Japanese is influenced by the tonal pattern and the particle following the word to be emphasized [6]. Therefore, in Japanese English, prominence in the Japanese language is negatively transferred to the English patterns, and the words for prominence are often inappropriately selected.

Some issues toward practical use of this approach need further investigations. One of them is to develop an automatic procedure for determining optimal values of a and b in equation in subsection 4.4.

We can address future issues concerning the irregularities of prominence from an individual syllable, specifically characteristics that involve the alignment of rising or falling pitch movements in Japanese English. Also, further research should verify the entire set of relevant correlates, that is, the pitch, duration and intensity should be included. Also, the effect of vowel realization, i.e., vowel quality, vowel epenthesis or vowel reduction, of Japanese English should be clarified. Extending the investigation to other cases, such as Japanese speakers of French or Chinese, will be helpful for exploring general pitch patterns of Japanese speaking a second language.

6. Concluding Remarks

This study has described the difference between Japanese English sentences and English speakers sentences. First, Japanese English and English speakers were comparatively analyzed. Then, irregularities of pitch patterns in sample Japanese English were acoustically modified. Finally, the modified speech was evaluated in a listening experiment taken by English speakers. The results of the experiments indicated practical verification of validity of the rules.

As mentioned in the Introduction, previous general studies on related themes adopted a bottom-up approach for extracting the difference. However, the results found in those studies were often found to be inaccurate when confirmed by re-synthesis of speech.

Irregularities in second language speech stems from a combination of several causes. It is generally difficult to extract definite cues as to what those features are by applying a general extraction method. Therefore, we employed an analysis framework where the characteristics that differentiated two language systems were predicted in advance. In addition, the effect of extracted prosodic features was verified by a listening test using speech re-synthesis. We consider that

this framework will make it possible to find the meaningful prosodic features of Japanese English.

References

[1] N. Minematsu, Y. Tomiyama, K. Yoshimoto, K. Shimizu, S. Nakagawa, and M. Dantsuji, "Development of English speech database read by Japanese to support CALL research," Proc. Int. Cong. Acoustics, pp.557–560, 2004.
[2] K. Imoto, Y. Tsubota, A. Raux, T. Kawahara, and M. Dantsuji, "Modeling and automatic detection of English sentence stress for computer-assisted English prosody learning system," Proc. Int. Conf. Spoken Language Processiong, pp.749–752, 2002.
[3] Y. Yamashita, K. Kato, and K. Nozawa, "Automatic scoring for prosodic proficiency of English sentences spoken by Japanese based on utterance comparison," IEICE Trans. Inf. & Syst., vol.E88-D, no.3, pp.496–501, March 2005.
[4] K. Imai, Eigo no tsukai kata, In Take of English series, no.4, Taishukan shoten, Tokyo, 1995. (in Japanese)
[5] K. Watanabe, Eigo no rhythm, intonation no shido, Taishukan shoten, Tokyo, 1994. (in Japanese)
[6] M. Sugito, Nihonjin no eigo, Nihongo onsei no kenkyu, no.2, Izumi shoin, Tokyo, 1996. (in Japanese)
[7] 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.
[8] T. Nariai, K. Tanaka, and Y. Itoh, "Comparative study of focal lengthening in the speech of native speakers and Japanese speakers of English," J. Acoust. Soc. T, vol.32, no.2, pp.54–61, 2011.
[9] O. Fujimura, Onsei lagaku genron, Iwanami Publishers, Tokyo, 2007, pp.164–167. (in Japanese)
[10] http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html
[11] J.B. Pierrehumbert and M.E. Beckman, "Japanese tone structure," Linguistic inquiry monograph, vol.15, MIT Press, Cambridge, 1988.
[12] D.R. Ladd, "Declination reset and the hierarchical organization of utterance," J. Acoust. Soc. Am., vol.84, no.2, pp.530–544, 1988.
[13] H. Kawahara, I. Masuda-Katsuse, and A. Cheveigne, "Restructuring speech representations using pitch adaptive pitch frequency smoothing and instantaneous-frequency-based F0 extraction," Speech Commun, vol.27, pp.187–207, 1999.
[14] 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 Processiong, pp.529–532, 2002.
[15] U. Weinreich, Language in contact: Findings and problems, Publications of the linguistic circle of New York, New York, 1953.
[16] J.E. Flege, M.J. Munro, and I.R.A. Mackay, "Factors affecting strength of perceived foreign accent in a second langage," J. Acoust. Soc. Am., vol.97, pp.3125–3134, 1995.
[17] K. Lambrecht, Information structure and sentence form: topic, focus, and the mental representations of discourse referents, Cambridge University Press, Cambridge, 1994.
[18] S. Takeda and A. Ichikawa, "Analysis of prominence in spoken Japanese sentences and application to text-to-speech synthesis," Speech Commun., vol.14, pp.171–196, 1994.

Tomoko Nariai is a Postdoctoral Fellow at the University of Tsukuba, Tsukuba, Japan. She graduated from the College of Humanities at the University of Tsukuba, receiving a Bachelor of Arts in Humanities degree in 2004. She completed the Master's Program and Doctoral Program in Library, Information and Media Studies at the University of Tsukuba, receiving her Master's and Ph.D. degrees in Information Science in 2007 and 2011, respectively.

Kazuyo Tanaka has been a professor at the University of Tsukuba, Tsukuba, Japan, since 2002. He received the B.E. degree from Yokohama National University, Yokohama, Japan, in 1970, and Dr. Eng. degree from Tohoku University, Sendai, Japan, in 1984. He was a research officer at Electrotechnical Laboratory (ETL), Tsukuba, Japan, from 1971 to 2000, where he was working on speech analysis, synthesis, and recognition, and also served as chief of the Speech Processing Section. In 2001, he was a leader of the Speech and Auditory Signal Processing Research Group at the National Institute of Advanced Science and Technology (AIST), Tsukuba, Japan. From 2005 to 2006, he served as a chairperson of the Speech Committee of the IEICE. His current interests include digital signal processing, spoken document processing, and human information processing. He is a member of IEEE, ISCA, Acoustical Society of Japan, and Japan Society of Artificial Intelligence.