Analysis of Japanese learners' English mispronunciation characteristics aiming at their objective evaluation

Hajime TSUBAKI[1,3], Takeru OGAMI[4], Mariko KONDO[2,3] and Yoshinori SAGISAKA[1,3]
[1] GITI, [2]SILS & [3]Language and Speech Science Research Laboratories, Waseda University, Japan
[4]TOYOTA MOTOR CORPORATION, Japan

Abstract

Aiming at objective evaluation of L2 (second language) learners' proficiency, we statistically analyzed mispronunciation characteristics of Japanese learners' English speech based on phonetic knowledge. Through automatic forced alignment using a speech recognition tool kit, phone deletion, substitution and insertion were detected by substituting the pronunciation dictionary reflecting learners' phonetic error characteristics. For phonetic knowledge, typical mispronunciations, vowel insertion, vowel change and consonant omission and change were taken into account. Mispronunciation analysis using 50 Japanese speakers showed consistent error characteristics along their subjective naturalness scores. The correlation sore 0.595 was obtained between subjective evaluation scores and objective measures by a multiple linear regression analysis, using the mispronunciation statistics as parameters.

1. Introduction

We have been studying L2 learner's speech characteristics aiming at automatic evaluation of them. The automatic evaluation of spoken language proficiency is expected to be useful becasue of well-defined common objective evaluation criteria and independence to variations of human raters' skill. Through the studies on automatic evaluation of L2 learners' proficiency, we expect that we can understand the important points of evaluation clearly and that they reasonably reduce load of teachers in language education.

Many studies have been conducted to enable objective evaluation of language fluency using spoken language information processing technology. In speech information processing, there are research focusing on L2 spectrum characteristics [1]-[3], and L2 duration control characteristics [4]-[6].

In this research, we analyzed learner's mispronunciation characteristics based on phonetic knowledge in English education and tried the objective evaluation of Japanese learners' English speech using mispronunciation characteristics In conventional studies, statistical likelihood measures of phone models in recognition have been mainly used to directly calculate acoustic distances. Whilst, the current study tried to find pronunciation errors directly using a newly proposed pronunciation dictionary reflecting learners' characteristics which could be used as information directly linked to L2 educational.

Comment [a1]: 英文を直しましたが、Abstr
＝要約中には書かない内容です。従って削除

For mispronunciation detection, phone alignment based on Hidden Markov Model was used. Phone sequence patterns reflecting possible mispronunciations were registered as phone sequences in a word dictionary. Using this mispronunciation dictionary, phone alignment was carried out using Japanese learners' English speech, the numbers of mispronounced phoneme sequence patterns were counted as learners' characteristics. The correlations between the observed mispronunciations and subjective evaluation scores given by specialists in English education were analyzed. We adopted the pronunciation errors showing high correlation to the subjective evaluation values, as parameters to estimate subjective scores using a multiple linear regression model.

2. Analysis on Japanese learners' English mispronunciation characteristics

Each language has its own phonemic system. In phonetics, it is well known that phone category denoted by [ ] is different from phoneme category denoted by / /. For example, both [s] and [ʃ] belong to /s/ phoneme category in Japanese, while the former belongs to /s/ and the latter belongs to /ʃ/. In this research, we focused on utterance characteristics based on the difference of the phonemic system and accent between Japanese and English in the Japanese L2 English learners' pronunciations. These utterance characteristics are divided into three major categories as follows.

Comment [a2]: 冗長な内容で何を述べたい
がわかりません。以下の内容の説明に繋げた
のだと理解し、簡単に補うことにします。

2.1. L2 phonetic changes considered in this study

In this study, three types of mispronunciations commonly observed in Japanese English were treated as the first step. They are vowel insertion, vowel substitution and consonant omission and substitution.

(1) Vowel insertion

Japanese is an open syllable language. From this phonemic constraint in Japanese L2 English learners' speech, a vowel is inserted at the end of syllable final consonant such as /beddo/ for /bed/(bed) or /a.teN.pu.to/ for /ə .tempt/(attempt).

(2) Vowel substitution

There is no distinction of stressed and unstressed vowel in Japanese. Japanese learners’ utterance of English short vowel can be changed to long vowel or diphthong. For example, /bʌt/ (*but*) in General American English may become /bat/ in Japanese English. The phoneme /ʌ/ is replaced to the phoneme /a/.

(3) Consonant omission and substitution

Since Japanese has fewer consonantal phonemes than English, Japanese learners omit unfamiliar English phonemes or map the phonemes to similar Japanese phonemes. For example, /nɔ□:rθ/ (*north*) in General American English may become /no:θ/ or /no:s/ in Japanese English. The phoneme /r/ is omitted and the phoneme /θ/ is replaced to the phoneme /s/.

2.2. Phone alignment using Japanese L2 learners' mispronunciation characteristics

Mispronunciation candidate phone sequences were generated for each word entry using above-mentioned three types of variations for phone alignment. These candidate phone sequences were registered in a word dictionary (Table 1). Phone alignment is carried out to allocate phone boundaries in speech signal by comparing acoustic similarities among candidate phone sequences using likelihood scores provided by Hidden Markov Model (HMM) models.

Table.1 Mispronunciation candidatephone ~~patterns~~sequences for word *blew*

Mispronunciation candidates	ARPABET phonetic transcription
[blu:]	b l uw
[bʊlu：]	b uh l uw
[bɹu：]	b r uw
[bʊɾu：]	b uh r uw

We trained acoustic model for the automatic alignment using TIMIT speech database as training data and HMM parameters as follows:

Training data: WSJ SI-285 database (speech
corpus) consisting of 37516 sentences
in total uttered by 284 speakers'
utterances (native speakers of
American English) and labeled by
ARPABET

HMM model: speaker-independent 3 state
monophone-model

Acoustic parameters: 39 parameters consisting of 12 dimensional
MFCC, log power, and theirΔs and ΔΔs

The alignment is performed based on the HMM model, and output is given as phone sequences in ARPABET [7].

3. Analysis of Japanese English learners' mispronunciations and their objective evaluation

From alignment results of Japanese L2 English utterance, mispronunciation phone sequences were extracted and summed as statistics. We analyzed correspondence relation between the statistics and subjective evaluation values. Using the statistics showing high correlation with human rating, we tried to estimate the subjective evaluation values.

3.1. Analysis data

We used 50 samples of Japanese L2 English speech selected from the Asian English voice corpus (AESOP: Asian English Speech cOrpus Project) [8].

Utterance person: Japanese native university student
Speech data: "North wind and sun" read
speech (all the 113 words)

3.2. The subjective evaluation value

7 specialists in English phonetics and education (5 Japanese raters and 2 English native raters) gave subjectivity evaluation values to the 50 data based on the following standard.

1) Evaluation criterion:
Native-likeliness of speaking

2) Evaluation category:
5 major and 4 minor scales
1-Very poor (Unintelligible) | 1.5 |
2-Poor | 2.5 | 3-Medium | 3.5 |
4-Good | 4.5 | 5-Very good (Native-like)

3.3. Correspondent analysis between the statistics and the subjectivity evaluation values

77 words of 113 words in the speech were found Japanese L2 English characteristics. The maximum number per learner was 77 words and the minimum

Formatted Table

Comment [a3]: アラインメント過程を示す
に見えないのですが、、、

number per learner was 50 words. The average was 50 words. Next, the correspondence relation between the statistics and the subjectivity evaluation values was investigated. We supposed that the fewer characteristics a learner had, the higher the learner's score was. Figure 1 is a scatter diagram which makes the vertical axis total number of the phoneme sequence pattern variants per learner, and makes the horizontal axis the learners' subjective evaluation value (fluency level). The correlation coefficient between the total number and the learners' subjective evaluation value is -0.454 and shows a tendency that the fewer total number a learner has, the higher the learner's score is.

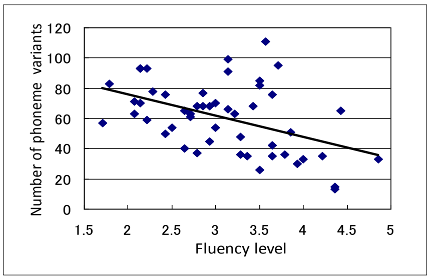

Figure. ~~2~~1 Correlation between phoneme sequence pattern variants and fluency level

3.3.1. The tendency of vowel insertion

The correlation coefficient between the total number of vowel insertion and the subjectivity evaluation values is -0.119. The highest correlation value is -0.204 for the vowel insertion /ʊ/. Since these kinds of vowel insertion occur in spite of the learners' proficiency level, it is not easy for Japanese L2 English learners to correct these pronunciation errors (Figures. 2 and 3).

3.3.2. The tendency of vowel substitution

The correlation coefficient between the total number of vowel change and the subjectivity evaluation values is -0.435 (Figure. 4). Based on the statistics, these kinds of vowel change occur across the levels. There is a tendency that these pronunciation errors are corrected as learners' proficiency level goes up (Figure. 5).

3.3.3. The tendency of consonant omission and substitution

The correlation coefficient between the total number of consonant omission and substitution, and the subjectivity evaluation values are -0.578 (Figure. 6). Based on the statistics, these kinds of consonant omission and change occur across the levels. There is a tendency that these pronunciation errors are corrected as learners' proficiency level goes up (Figure. 7).

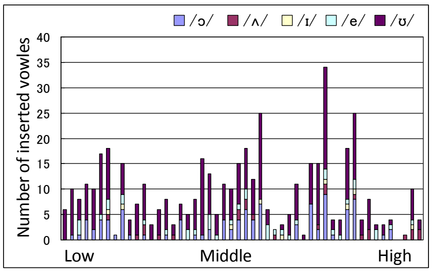

Figure. 2 Relation between vowel insertion and fluency level

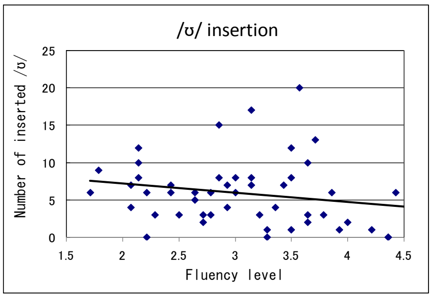

Figure. 3 Relation between vowel /ʊ/ insertion and fluency level

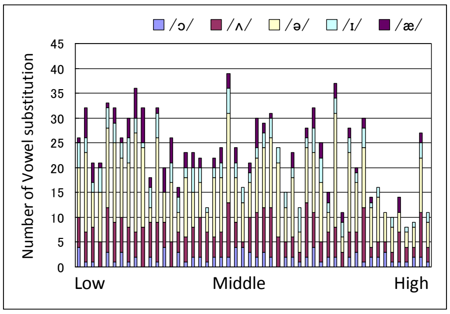

Figure. 4 Relation between vowel substitution and fluency level

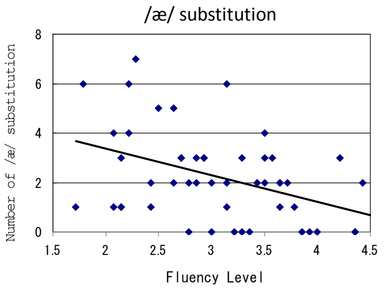

Figure. 5 Relation between vowel /æ/ substitution and fluency level

3.4. Estimation experiment of subjective evaluation value

Using the statistics showing higher correlation with the subjective evaluation values as parameters, we tried to estimate the values. Considering multicollinearity among the statistics, we chose the numbers of vowel change /ə/, /æ/ and /ɪ/, consonant omission /r/, and consonant change /θ/.

50 Japanese L2 English utterance data were divided into 40 data as training data and 10 data as test data. Using the actual subjective evaluation values as parameters, we made multiple regression formula for subjective evaluation value estimation. We obtained estimated subjective evaluation values and correlation coefficient 0.595 between the actual and estimated values (Figure.8).

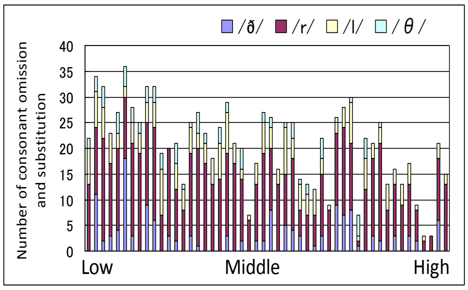

Figure. 6 Relation between consonant omission and substitution, and fluency level

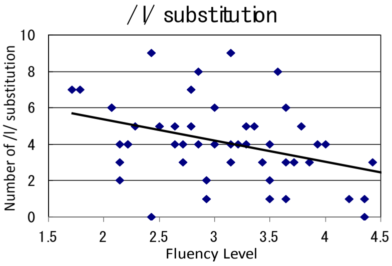

Figure. 7 Relation between / l / substitution and fluency level

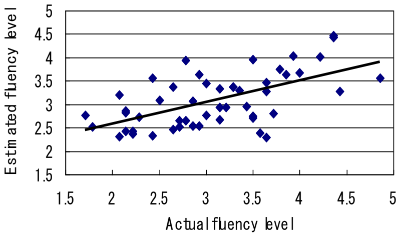

Figure. 8 Correlation between actual and estimated subjective evaluation

4. Summary and future problems

To evaluate the Japanese L2 English learner's speech proficiency, we examined the applicability of the phoneme patterns reflecting their utterance characteristics extracted from their speech. Using the multiple regression analysis, the speech proficiency (subjective evaluation value) was estimated by the statistical phoneme patterns. The estimated values showed a high correlation 0.595 to the actual values.

These results indicate the usefulness of the statistical phoneme patterns for the estimation of L2 English speech proficiency. As it is confirmed that the statistical phoneme patterns work for L2 proficiency evaluation, we would like to use other statistics such as duration and F0 control characteristics.

References

[1] K.Hirabayashi, S.Nakagawa. Automatic evaluation of English pronunciation by Japanese speakers using various acoustic features and pattern recognition techniques. Proc. Interspeech, pp.598-601, 2010

[2] H.Wang, C.J.Waple, T.Kawahara. Computer assisted language learning system based on dynamic question generation and error prediction for automatic speech recognition. J. Speech Communication, Vol.51, No.10, pp.995-1005, 2009.

[3] H.Wang, T.Kawahara. Effective prediction of errors by non-native speakers using decision tree for speech recognition-based CALL system. IEICE Trans., Vol.E92-D, No.12, pp.2462-2468, 2009.

[4] Nakamura, S., Tsubaki, H., Kondo, Y., Nakano, M., & Sagisaka, Y. 2007. Tempo-normalized measurement and test set dependency in objective evaluation of English learners' timing characteristic. Proceedings of ICPhS 2007, 1733-1736.

[5] S. Nakamura, S. Matsuda, H. Kato, M. Tsuzaki and Y. Sagisaka, “Objective evaluation of English learners' timing control based on a measure reflecting perceptual

characteristics", Proc. IEEE ICASSP, pp.4837-4840, 2009. 4.

[6] S. Nakamura, H. Kato and Y. Sagisaka, "Effects of Mora-timing in English Rhythm Control by Japanese Learners", Proc. INTERSPEECH 2009 pp.1539-1542, 2009. 9.

[7] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, V. Zue The DARPA TIMIT acoustic-phonetic continuous speech corpus Published in 1992

[8] Tanya Visceglia, Chiu-yu Tseng, Mariko Kondo, Helen Meng and Yoshinori Sagisaka Phonetic Aspects of Content Design in AESOP (Asian English Speech cOrpus Project) Oriental-COCOSDA, 2009