Effectiveness of Linguistic and Learner Features to Listenability Measurement Using a Decision Tree Classifier

Katsunori Kotani
Kansai Gaidai University
Hirakata, Osaka, Japan
kkotani@kansaigaidai.ac.jp

Takehiko Yoshimi
Ryukoku University
Otsu, Shiga, Japan
yoshimi@rins.ryukoku.ac.jp

Abstract

In learning Asian languages, learners encounter the problem of character types that are different from those in their first language, for instance, between Chinese characters and the Latin alphabet. This problem also affects listening because learners reconstruct letters from speech sounds. Hence, special attention should be paid to listening practice for learners of Asian languages. However, to our knowledge, few studies have evaluated the ease of listening comprehension (listenability) in Asian languages. Therefore, as a pilot study of listenability in Asian languages, we developed a measurement method for learners of English in order to examine the discriminability of linguistic and learner features. The results showed that the accuracy of our method outperformed a simple majority vote, which suggests that a combination of linguistic and learner features should be used to measure listenability in Asian languages as well as in English.

1 Introduction

An important task of language teachers is to choose reading/listening materials appropriate for the proficiency of their learners so as to prevent decreases in learning motivation. However, this task can be a heavy burden for language teachers when they introduce computer-assisted language learning/teaching (CALL/T) techniques. Although CALL/T allows language teachers to use different reading/listening materials for each learner, it also increases the number of materials that they must evaluate for appropriateness. To address this issue, alternative methods that automatically measure the ease of reading comprehension (readability) have been developed.

However, although the majority of previous studies have addressed the measurement of readability: Japanese by Sato et al. (2008); Chinese by Sung et al. (2015), among others, they have not addressed the ease of listening comprehension (henceforth, listenability). Several studies have examined listenability for English learners (Kiyokawa 1990; Kotani et al. 2014; Kotani & Yoshimi 2016; Yoon et al. 2016); however, to the best of our knowledge, no previous studies on listenability for learners of Asian languages such as Chinese, Korean, and Japanese have been conducted.

The method of Kiyokawa (1990) measured listenability based on the length of sentences and the difficulty of words. It was hypothesized that the listenability of a sentence decreases as it becomes longer and contains more advanced vocabulary words. Kotani et al. (2014) suggested the possibility of using different linguistic elements such as phonological features, and addressed this question by measuring listenability based on various linguistic features, including speech rate and the frequency of phonological modification patterns such as linking. In addition, their method used listening test scores as a learner feature to measure listenability relative to proficiency. This is because sentences with less listenability for learners at the beginner level might be easy for those at the advanced level. However, because that study focused on the accuracy of measurement, the question of discriminability of

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/

Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications, pages 6–10, Osaka, Japan, December 12 2016.

linguistic and learner features for the measurement of listenability remained. The discriminability of linguistic features was examined by Yoon et al. (2016), who used multiple regression analysis to measure listenability; however, they did not examine the discriminability of learner features. Hence, the discriminability of both linguistic and learner features still has yet to be examined.

Given this background, the purpose of this study was to attempt to answer the following two research questions by measuring listenability on the basis of linguistic and learner features:

(1) How accurately can listenability be measured using linguistic and learner features?
(2) Which of linguistic and learner features are discriminative for the measurement of listenability?

To answer these questions, in this study, we developed a listenability measurement method using a decision tree classification algorithm (Quinlan 1992) that classifies sentences into five levels of listenability in order to determine the accuracy of listenability measurement and the discriminability of linguistic and learner features to this classification. Although the linguistic and learner features examined were effective for listenability measurement in English, they were not English-specific, which suggests that they may also be useful for the measurement of listenability in Asian languages.

2 Linguistic and Learner Features

Listenability is measured based on linguistic and learner features. Linguistic features explain the difficulty of a sentence, and learner features explain the proficiency of a learner. The linguistic (Chall 1948; Fang 1966; Kiyokawa 1990; Messerklinger 2006; Kotani et al. 2014; Kotani & Yoshimi 2016; Yoon et al. 2016), and learner features (Kotani et al. 2014; Kotani & Yoshimi 2016) used in this study were originally described elsewhere.

Linguistic features consist of sentence length, mean word length, multiple syllable words, word difficulty, speech rate, and phonological modification patterns. Sentence length is calculated based on the number of words in a sentence. Mean word length is derived from the mean number of syllables per word. Multiple syllable words refer to the number of multiple syllable words in a sentence. Word difficulty is derived from the rate of words absent from Kiyokawa's basic vocabulary list for words in a sentence. Speech rate is calculated in terms of spoken words per minute. Phonological modification patterns are derived from the rate of phonologically modified words in a sentence. The types of phonological modification patterns are: elision (elimination of phonemes), in which vowel sounds immediately follow a stressed syllable, such as the second "o" sound in "chocolate"; reduction (weakening a sound by changing a vowel to a schwa), such as vowel sounds in personal/interrogative pronouns, auxiliaries, modals, prepositions, articles, and conjunctions; contraction (combining word pairs), such as a modal with a subject noun; linkage (connecting final and initial word sounds), such as connected a word ending with an "n" or "r" sound with a word starting with a vowel sound, for example, "in an hour" and "after all"; and deduction (elimination of sounds between words), in which words share the same sound, for example, "good day".

Learner features consist of listening test scores, learning experience, visiting experience, and listening frequency. Listening test score refers to scores on the Test of English for International Communication (TOEIC). Learning experience refers to the number of months for which learners have been studying English. Visiting experience refers to the number of months learners have spent in English-speaking countries. Listening frequency refers to scores on a five-point Likert scale for the frequency of English use (1: infrequently, 2: somewhat infrequently, 3: moderate, 4: somewhat frequently, and 5: frequently).

3 Training/Test Data

Training/test data for a decision tree classification algorithm were constructed using the learner corpus of Kotani et al. (2014), which includes learners' judgment of listenability. Listenability was judged by learners of English as a foreign language using scores on a five-point Likert scale (1: easy, 2: somewhat easy, 3: average, 4: somewhat difficult, or 5: difficult). Scores were judged on a sentence-by-sentence basis where each learner listened to and assigned scores for 80 sentences from four news clips selected from the editorial and special sections for English learners on the Voice of America (VOA) website (http://www.voanews.com). News clips in the special section were intended for learners, while news clips in the editorial section were intended for native speakers of English. The news clips in the special section consisted of short, simple sentences using the VOA's basic vocabulary of 1,500 words; idiomatic expressions were avoided. By contrast, the news clips in the editorial section were made without any restrictions on vocabulary and sentence construction, as long as they were appropriate as news clips for native speakers of English. The speech rate of the news clips in the special section were two-thirds slower than those in the editorial section, which were read aloud at a natural speech rate of approximately 250 syllables per minute (Robb & Gillon 2007).

The learners were 90 university students (48 males, 42 females; mean age ± SD, 21.5 ± 2.6 years) who were compensated for their participation. All learners were asked to submit valid scores from TOEIC tests taken in the current or previous year. The mean TOEIC listening score was 334.78 ± 98.14. The minimum sore was 130 (n = 1), and the maximum score was 495 (n = 8).

Although the training/test data should have consisted of 7,200 instances (90 learners × 80 sentences) for valid listenability measurement, only 6,804 instances were actually observed. Assuming that the missing 396 instances resulted from listening difficulties, these instances were scored as having the lowest listenability. Most instances (25.2%) were scored in the middle range (3) of listenability, and the fewest instances (15.8%) were scored in the high range (2). Listenability scores of 1, 4, and 5 were given by 21.7%, 20.8%, and 16.5% of the learners, respectively.

Table 1 shows the means and SDs of the linguistic and learner features in the training/test data.

Table 1. Descriptive statistics of linguistic (n = 80) and learner features (n = 90)

Type	Feature		Mean	SD
Linguistic	Sentence length		17.6 (words)	7.5
	Mean word length		1.7 (syllables)	0.3
	Multiple syllables		11.2 (words)	7.0
	Difficult words		0.4 (words)	0.1
	Speech rate		199.3 (words per minute)	49.2
	Phonological modification pattern	Elision	0.0 (points)	0.1
		Reduction	0.2 (points)	0.2
		Contraction	0.1 (points)	0.1
		Linking	0.0 (points)	0.0
		Deduction	0.4 (points)	0.2
Learner	Listening test score		334.8 (points)	97.6
	Learning experience		123.2 (months)	36.6
	Visiting experience		11.3 (months)	25.8
	Listening frequency		2.1 (score)	1.1

4 Experiment

Listenability was measured on the basis of linguistic and learner features using a decision tree classification algorithm implemented on C4.5 software (Quinlan 1992). All settings were taken as defaults, and classification was evaluated using five-fold cross validation.

Table 2. Confusion matrix for the test data

Method's / Learner's	Listenability 1	Listenability 2	Listenability 3	Listenability 4	Listenability 5
Listenability 1	1116 (71.4%)	190	169	46	42
Listenability 2	299	293 (25.8%)	348	125	70
Listenability 3	188	307	740 (40.8%)	439	139
Listenability 4	72	161	463	574 (38.3%)	228
Listenability 5	78	59	146	247	661 (55.5%)

The results of the five-fold cross validation tests, as well as the confusion matrix for the test data, where the rows indicate the correct classification and the columns indicate the selected classes, are shown in Table 2. The accuracy of classification rate was 47.0% ((1116+293+740+574+661)/7200) in the test data. Although this might be insufficient for validating our listenability measurement method, we believe that the method can still be judged as valid through a comparison with the accuracy attained by a simple majority vote (25.2%) as a baseline.

We calculated the accuracy for each listenability score from 1 to 5, which is shown as bracketed numbers in Table 2. The accuracy varied from 25.8% (293/(299+293+348+125+70)) to 71.4% (1116/(1116+190+169+46+42)). As this examination was not conclusive, it remains for the future study to examine why the method showed the different accuracies in more detail.

Using the five-fold cross validation test, five decision trees (I–V) were generated. In four of the five decision trees, the same type of linguistic and learner features were allocated at the root nodes, the first-level child nodes (child nodes originating from the root nodes), and the second-level child nodes (child nodes originating from the first-level child nodes). Parts of the decision tree (I–IV) can be seen in Figure 1; the different roots (V) are shown in bold. Part V of the decision tree is shown in Figure 2.

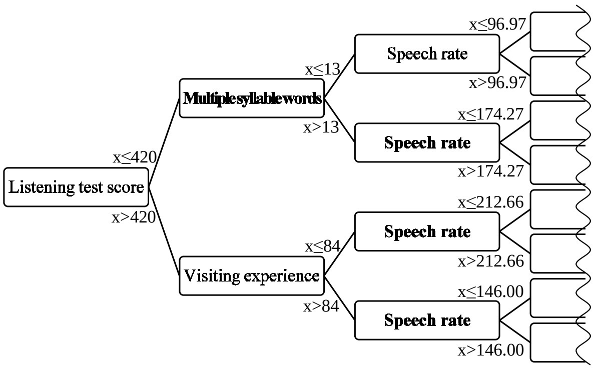

Fig. 1. Decision tree (I–IV)

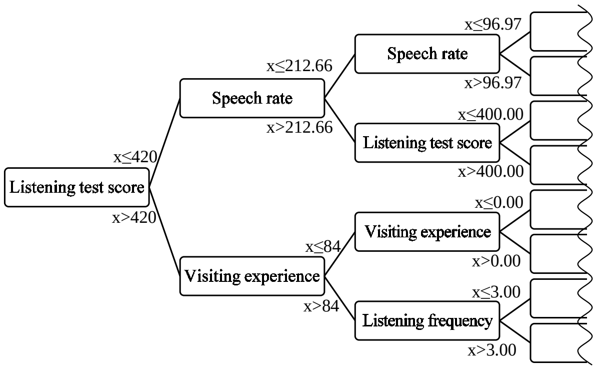

Fig. 2. Decision tree (V)

As the listening test score was allocated at the root node of the five decision trees, this feature was regarded as the most discriminative. Visiting experience was allocated at the first-level child node of the decision trees, and was therefore judged as the second most discriminative feature. The third most discriminative feature was regarded as the speech rate, because it was allocated at either the first- or second-level child node in each tree.

5 Conclusion

In this study, we examined the measurement of listenability for learners of English as a foreign language. We found that learner features were discriminative for the measurement accuracy. This finding suggests that learner features should be taken into account when measuring listenability for learners of Asian languages.

Although the accuracy was not high, our method outperformed a simple majority vote. In the future, using this method as a baseline, we plan on developing a listenability measurement method for Asian languages that would outperform that for English.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers, 22300299, 15H02940.

Reference

Chall, J. S. & Dial, H. E. 1948. Listener Understanding and Interest in Newscasts. *Educational Research Bulletin*, 27(6): 141–153+168.

Fang, I.E. 1966. The Easy Listening Formula. *Journal of Broadcasting & Electronic Media*, 11(1): 63–68.

Kiyokawa, H. 1990. A Formula for Predicting Listenability: the Listenability of English Language Materials 2. *Wayo Women's University Language and Literature*, 24: 57–74.

Kotani, K., Ueda, S., Yoshimi, T., & Nanjo, H. 2014. A Listenability Measuring Method for an Adaptive Computer-assisted Language Learning and Teaching System. *Proceedings of the 28th Pacific Asia Conference on Language, Information, and Computation*: 387–394.

Kotani, K. & Yoshimi, T. 2016 (in press). Learner Feature Variation in Measuring the Listenability for Learners of English as a Foreign Language. *Proceedings of the 1st International Workshop on Emerging Technologies for Language Learning (ETLL 2016)*.

Messerklinger, J. 2006. Listenability. *Center for English Language Education Journal*, 14: 56–70.

Quinlan, J. Ross. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1992.

Robb, M. P. & Gillon, G. T. 2007. Speech Rates of New Zealand English- and American English-speaking Children. *Advances in Speech-Language Pathology*, 9(2): 1–8.

Sato, S., Matsuyoshi, S., & Kondoh, Y. 2008. Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*: 654-660.

Sung, Y. T., Chen, J. L., Cha, J. H., Tseng, H. C., Chang, T. H., & Chang, K. E. 2015. Constructing and Validating Readability Models: the Method of Integrating Multilevel Linguistic Features with Machine Learning. *Behavior Research Methods*, 47(2): 340-354.

Yoon, S-Y., Cho, Y., & Napolitano, D. 2016. Spoken Text Difficulty Estimation Using Linguistic Features. *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*: 267–276.