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

Katsunori Kotani* and Takehiko Yoshimi**

(Received 30 November 2016 and accepted in revised form 20 June 2017)

Abstract As the ease of grasping the contents of listening material influences learners' motivation and learning outcome, language teachers need to choose materials appropriate for the proficiency of their learners. This heavy task has been addressed by using a traditional readability measurement method to develop an automatic measurement method of the ease of listening comprehension using linear regression analysis for listening materials. Because machine learning such as decision tree classification can properly handle different types of features, recent readability measurement methods use classification approaches such as a decision tree. Then, we proposed a measurement method using decision tree classification for linguistic features of listening materials as well as learner features of listening proficiency. The experimental results showed that the accuracy of our method (47.0%) was better than the baseline accuracy (25.2%), and that the listening test score and visiting experience in English speaking areas among the learner features were discriminative for the measurement accuracy.

Keywords: computer-assisted language learning/teaching, listenability, decision tree, linguistic feature, learner feature

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, they have not addressed the ease of listening comprehension (henceforth, we call this listenability). Several studies have examined listenability for English learners^(1–4).

The method of Kiyokawa⁽¹⁾ measured listenability based on the length of sentences and the difficulty of words. Kiyokawa hypothesized that the listenability of a sentence decreases as it becomes longer and contains

more advanced vocabulary. Kotani et al.⁽²⁾ and Kotani and Yoshimi⁽³⁾ 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 low listenability for learners at the beginner level might be easy for those at the advanced level. However, because their method focused on the accuracy of measurement, the question of being able to discriminate between linguistic and learner features for the measurement of listenability remained. The discriminability of linguistic features was examined by Yoon et al.⁽⁴⁾, 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, our purpose in 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 the linguistic and learner features are discriminative for the measurement of listenability?

*School of International Professional Development, Kansai Gaidai University, Japan

**Department of Media Informatics, Faculty of Science and Technology, Ryukoku University, Japan

To answer these questions, we developed a listenability measurement method using a decision tree classification algorithm⁽⁵⁾ 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 in this classification. A decision tree classification algorithm was chosen because we placed more importance on the interpretability of classification results than on classification accuracy. A decision tree classification algorithm is capable of generating a more interpretable decision tree, although the classification accuracy is lower than other well-known classification algorithms such as random forests and multi-class support vector machines. The experimental results of our decision tree classification algorithm suggested that the linguistic and learner features examined were effective for listenability measurement.

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^(1-4, 6-8) and learner^(2, 3) features 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 conditions of phonological modification patterns are summarized in Table 1.

Learner features consist of listening test scores, learning experience, visiting experience, and listening frequency. Listening test scores refer 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.

Table 1. Conditions for Phonological Modification Patterns.

Type	Condition for phonological modification
Elision	(i) vowel sound immediately following a stressed syllable such as the second "o" sound in "chocolate" (ii) a consonant followed by similarly articulated sound such as (a) continuous same sound as in "unknown," (b) continuous plosive sound as in "c" sound and "t" sound of "doctor," and (c) plosive sound followed by nasal sound as in "suddenly"
Reduction	vowel sound in a functional word such as personal/interrogative pronouns, auxiliaries, modals, prepositions, articles, and conjunctions
Contraction	a pair of a be-/have-auxiliary, or a modal with a subject noun/an interrogative pronoun/a negative adverb "not"
Linking	(i) a word ending with "n" or "r" sound followed by a word starting with vowel sound as in "in an hour" and as in "after all" (ii) a word followed by an indefinite article, a preposition, or a conjunction
Deduction	(i) words sharing the same sound as in "good day" (ii) words sharing the same type of pronunciation such as plosive, affricative, fricative, nasal, or lateral sound as in "next chance"

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.⁽²⁾, 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

Table 2. Distribution of Listenability.

Listenability	Frequency (<i>n</i>)	Relative frequency (%)
1	1563	21.7
2	1135	15.8
3	1813	25.2
4	1498	20.8
5	1191	16.5

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⁽⁹⁾.

The learners were 90 university students (48 men, 42 women; mean age $\pm$ SD, 21.5 $\pm$ 2.6 years) who were financially compensated for their participation. All learners were asked to submit valid TOEIC scores for tests taken in the current or previous year. The mean TOEIC listening score was 334.8 $\pm$ 98.1. The minimum score 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 $\times$ 80 sentences) for valid listenability measurement, only 6,804 instances were valid. Assuming that the missing 396 instances resulted from listening difficulties, these instances were arbitrarily scored as having the lowest listenability. As Table 2 shows, 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 3 shows the means and SDs of the linguistic and learner features in the training/test data.

Table 3. Descriptive Statistics of Linguistic ($n=80$) and Learner ($n=90$) Features

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)
		Reduction	0.2 (points)
		Contraction	0.1 (points)
		Linking	0.0 (points)
		Deduction	0.4 (points)
Learner	Listening test score	334.8 (points)	98.1
	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⁽⁵⁾. All settings were taken as defaults, and the classification was evaluated using five-fold cross validation.

The results of the five-fold cross validation, 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 4.

EFFECTIVENESS OF LINGUISTIC AND LEARNER FEATURES FOR LISTENABILITY MEASUREMENT USING A DECISION TREE CLASSIFIER

Table 4. 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%)

Figure 1. Decision Trees (I–IV).

Figure 2. Decision Tree (V).

The accuracy of the 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 baseline accuracy (25.2%), which is defined as the percentage of examples in the most frequently occurring listenability class (listenability 3) in the training data (see Table 2).

We calculated the accuracy for each listenability score from 1 to 5, which is shown as bracketed numbers in Table 4. The accuracy varied from 25.8% (293/(299+293+348+125+70)) to 71.4% (1116/(1116+190+169+46+42)), depending on the listenability scores. As this examination was not conclusive, it remains for a future study to examine the relation of the accuracy with

the listenability scores in more detail.

Using the five-fold cross validation, we generated five decision trees (I–V). In four of the five decision trees, the same types 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 trees (I–IV) and (V) can be seen in Figure 1 and Figure 2, respectively. The different nodes between the decision trees (I–IV) and (V) are shown in bold.

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 discrimina-

tive 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.

The listening test score and visiting experience were more discriminative for the listenability measurement than learning experience and listening frequency, which was incompatible with the assumption that the examined learner features explain learners' listening proficiency. In a future study, we plan to examine different learner features, and explicate the contribution of learner features.

5. Conclusion

In this study, we examined the measurement of listenability for learners of English as a foreign language. Although the accuracy was not high, our method outperformed the baseline accuracy. 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 English learners.

In the future, using this listenability measurement method as a base, we plan on improving it for English learners. In addition, as TOEIC evaluates learners' listening proficiency featuring four accents of English, American, Canadian, British and Australian, we plan to develop a listenability measurement method on the basis of an English corpus featuring different accents.

Acknowledgement

We would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. All remaining inadequacies are ours alone. This work was supported by JSPS KAKENHI Grant No. 22300299 and No. 15H02940.

References

- (1) Kiyokawa, H.: "A Formula for Predicting Listenability: The Listenability of English Language Materials 2", Wayo Women's University Language and Literature, Vol. 24, pp. 57–74 (1990).
- (2) Kotani, K., Ueda, S., Yoshimi, T. et al.: "A Listenability Measuring Method for an Adaptive Computer-assisted Language Learning and Teaching System", Proc. of the 28th Pacific Asia Conference on Language, Information, and Computation, pp. 387–394 (2014).
- (3) Kotani, K. and Yoshimi, T.: "Learner Feature Variation in Measuring the Listenability for Learners of English as a Foreign Language", Proc. of the 1st International Workshop on Emerging Technologies for Language Learning (2016).
- (4) Yoon, S.-Y., Cho, Y. and Napolitano, D.: "Spoken Text Difficulty Estimation Using Linguistic Features", Proc. of the 11th Workshop on Innovative Use of NLP for Building Educational Applications, pp. 267–276 (2016).
- (5) Quinlan, J. R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1992).
- (6) Chall, J. S. and Dial, H. E.: "Predicting Listener Understanding and Interest in Newscasts", Educational Research Bulletin, Vol. 27, No. 6, pp. 141–153, p. 168 (1948).
- (7) Fang, I. E.: "The Easy Listening Formula", J. of Broadcasting & Electronic Media, Vol. 11, No. 1, pp. 63–68 (1966).
- (8) Messerklinger, J.: "Listenability", Center for English Language Education J., Vol. 14, pp. 56–70 (2006).
- (9) Robb, M. P. and Gillon, G. T.: "Speech Rates of New Zealand English- and American English-Speaking Children", Advances in Speech-Language Pathology, Vol. 9, No. 2, pp. 173–180 (2007).