Development of an automatic judging system for determining the difficulty levels of English audio materials

Hironobu Okazaki[1], Kanji Watanabe[2], Shinichi Hashimoto[3], Mitsuko Suzuki[4], Eri Fukuda[5], and Kazuhiko Kido[6]

Abstract. Okazaki and Nitta (2005) developed an e-learning program called PLIMA (your Personal LIstening Manager) which focuses on improving poor phonological analysis such as the inability to hear liaison or unstressed sounds. However, the fact remains that there are certain types of liaison sounds that learners do not catch, and some types they do. This means if we specify what type, in other words, what level of liaison sounds they cannot catch, that will help PLIMA offer more effective learning to learners who lack certain listening skills. To automatically determine the difficulty levels of the English audio materials, text evaluation techniques using lexical databases have been employed in many cases so far. However, automatic judging systems on the difficulty levels of the English audio materials themselves, not using lexical databases, are almost unseen. In this paper, we are going to propose a listening difficulty level determination system, to overcome such a challenging task. It applies a processing technique to extract the acoustic feature quantities that are employed in speech recognition engines.

Keywords: difficulty level determination, acoustic feature quantities, e-learning.

1. Akita Prefectural University; okazaki@akita-pu.ac.jp.
2. Akita Prefectural University; kwatanabe @akita-pu.ac.jp.
3. The University of Electro-Communications; heiwanian@gmail.com.
4. Soka High School; smitsuko1129@gmail.com.
5. Chugoku Junior College; efukuda@cjc.ac.jp.
6. International Pacific University; k.kido@ipu-japan.ac.jp.

How to cite this article: Okazaki, H., Watanabe, K., Hashimoto, S., Suzuki, M., Fukuda, E., & Kido, K. (2014). Development of an automatic judging system for determining the difficulty levels of English audio materials. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds), *CALL Design: Principles and Practice*; *Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands* (pp. 278-283). Dublin: Research-publishing.net. doi:10.14705/rpnet.2014.000231

1. Introduction

One of the major weaknesses of Japanese EFL learners is the lack of ability to correctly comprehend the simple daily conversations by native English speakers. It seems that Japanese learners rely heavily on a top-down approach to understanding the natural conversation, trying to piece together meaning from the few words they are able to discern and the overall direction of the content, and often give inappropriate responses because of miscomprehension. In order to minimize such reliance on guessing, it is necessary to increase learners' ability to understand content bottom-up by building up their knowledge of the English language.

Our previous research project (Okazaki & Nitta, 2005) developed an effective e-learning program called PLIMA (your Personal LIstening MAnager) that personalizes tasks so that learners can intensively work on their weakest areas to increase their knowledge of English and to increase their percentage of using bottom-up approaches to understanding.

Figure 1 is the data of 270 students from three Japanese universities, showing their listening ability when native English speakers speak at normal rates. The column chart shows word recognition ratios, which indicate what percent of words that students already know are recognizable. At the time of testing, the students already knew more than 99% of the words the speaker used. The line chart shows the tendency of students who try to guess the meaning of what was said (Okazaki & Nitta, 2005).

Figure 1. Word recognition and meaning guessing

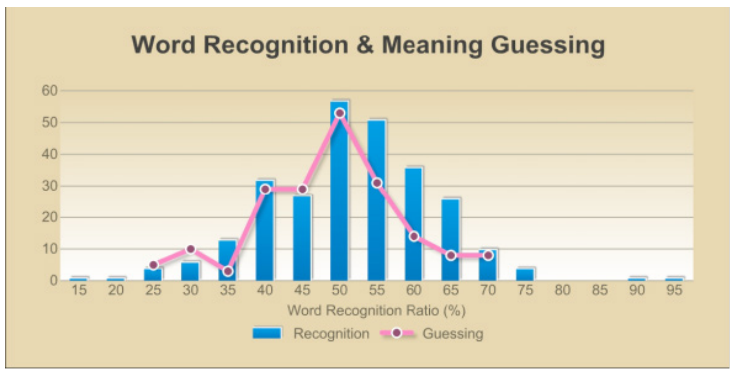


Figure 1 tells us that, when native speakers speak to Japanese EFL students in the same manner as they speak to other native speakers, average students cannot

comprehend half of the words spoken in spite of having knowledge of the words. Instead, they try to guess the meaning of what was said based on the half they were able to comprehend. As the word recognition ratio improves, they can get enough information and the guessing line declines. The line also declines as the word recognition ratio drops lower than the average. That is probably because students might not be able to comprehend enough to even guess, and just give up trying to understand the speakers.

PLIMA clearly helped learners overcome problems of not being able to hear liaison sounds and unstressed syllables, as well as increased their ability to analyse phonemes (Okazaki, Nitta, & Kido, 2011). However, it is true that there are certain types of liaison sounds that learners do not catch, and some types they do. That means if we specify what type, or, what level of liaison sounds they cannot catch, this will help PLIMA offer more effective learning to learners who lack certain listening skills.

The next challenge is to develop an automatic system to determine the difficulty levels of the English audio materials.

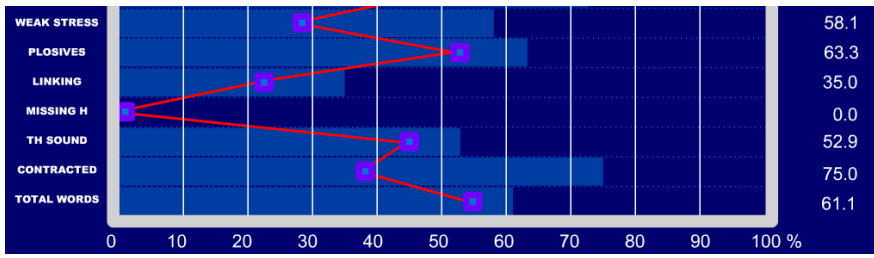
2. Method

2.1. Making use of PLIMA data

In PLIMA, there is a large volume of stored data from monitoring hundreds of college students practicing on this listening system, including the results of assessment tests they were required to take when using the system for the first time (see Figure 2). First of all, we analyzed the results of the assessment test taken by 83 Japanese college students that consists of 30 sentences derived from films and found some interesting and suggestive results. For example, the pronoun ‘you’ appears more than 10 times and the rate of miscomprehension is very broad: 10–70%. In the case of the word ‘what’, the word appears 4 times in 30 sentences and the rate of miscomprehension is 18.2%, 27.3%, 29.1% and 69.1%; one in 4 ‘what’s is apparently hard to comprehend for Japanese EFL learners.

These numbers demonstrate that it can be impossible to measure the difficulty levels of spoken English only through text-based analyses. In other words, if it is possible to find out the acoustic features when the percentages of correct answers are low, the priority to overcome will become more clear. Therefore, based on the result of analyses above, we started considering how to digitize the difficulty levels of spoken English by using the method of acoustic information processing.

Figure 2. An example of the results of the PLIMA assessment test



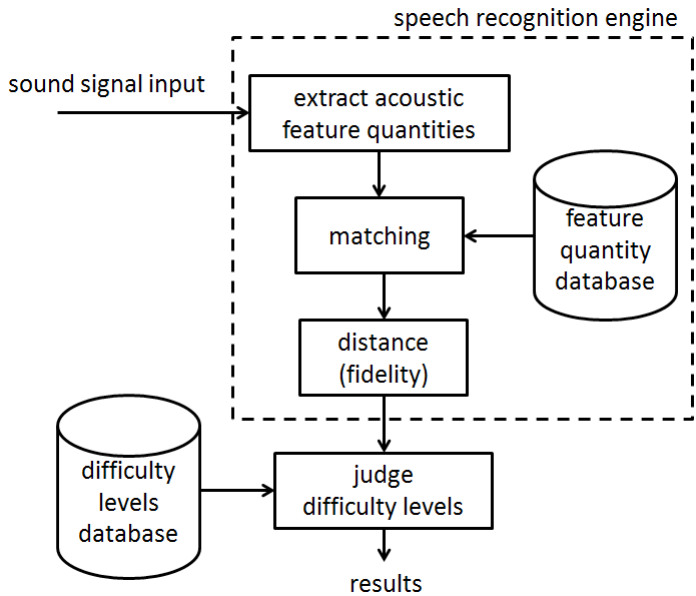
2.2. A model of an automatic judging system

Generally speaking, speech recognition systems use frequency analysis to process unknown sound signals by matching the features of that signal to known signal features which have been stored beforehand in a database. Matching is usually realized by comparing an aspect of the sound features, and then finding a counterpart which is "closest in distance" to that particular sample. Problems may arise when an input sound signal is difficult to recognize. The output from the database which is "closest in distance" may be using a very wide tolerance, or there may be a large number of candidates which are "closest in distance", making it difficult to make a clear determination of the match. Taking all these factors into consideration, however, it seemed plausible to configure software that would allow for determining the difficulty level of an English listening sample by comparing the acoustic sound features with items of known difficulty level.

Our goal is to design a practical automatic judging system to determine the difficulty of a listening passage by comparing distance measures of acoustic features inside a speech recognition engine. Figure 3 is a diagram of the model of that system. The portion enclosed by a dotted line represents existing speech recognition technology in the form of publicly available freeware. First, the acoustic features of the input are extracted using mel-frequency cepstral coefficients (MFCC), which is often used for speech recognition (Lee & Kawahara, 2009). This allows for the calculation of the power and first and second differences of the sound data. In general speech recognition processing, these features are compared to known items within a database and the match is output by pairing items that are closest in distance in a particular aspect of the acoustic feature. Our system then compares that item to items in a second database of items of known difficulty levels, and assigns a value to the output which indicates its difficulty level. As a technical detail, instead of using MFCC, a simple cepstrum (CEP) or other methods of

extracting the acoustic features may be used in the final version of the software. For the purpose of matching items, the Hidden Markov Model will most likely be used. The second database for determining difficulty level is concurrently under development.

Figure 3. A model of an automatic judging system



3. Our future system

In order to make use of a particular English listening material for class, such as listening from a broadly available news source, the instructor or the students themselves have to rely mainly on personal judgement to determine if that material is appropriate or not. Recently, the movement to incorporate corpus linguistics in the discussion for determining the difficulty level of language material is gaining popularity, but this discussion is strictly limited to text-based methods of analysis, and cannot be applied to analysing sound data. As this paper explains above, through this study, we hope to apply sound analysis engineering technology to the task of determining the difficulty level of listening material, which up to now has depended heavily on human judgement. If this technology can be successfully developed, it can be applied to e-learning programs. The automatic judging system that we are aiming to create will be part of a completely automatic e-learning program.

Acknowledgements. This research was partially supported by Grant-in-Aid for Scientific Research (C) (18520444) by Japan Society for the Promotion of Science (JSPS).

References

Lee, A., & Kawahara, T. (2009). Recent development of open-source speech recognition engine Julius. *The Proceedings of APSIPA ASC 2009* (pp. 131-137).

Okazaki, H., & Nitta, H. (2005). Internet-based personal listening program: A program for Japanese EFL learners to overcome individual weaknesses. In *FLEAT5 Proceedings* (pp. 92-99).

Okazaki, H., Nitta, H., & Kido, K. (2011). Practical use and validation of an original learning assistance program "Movie English Listening System". *ATEM Journal, 17*, 21-31.