2021 年度コンピュータサイエンス研究 III 最終審査会資料 2022 年 2 月 3 日

A method of accent and intonation correction during shadowing exercise for learners of Japanese

G2120006 シェリデショーワ・アンナ

1 Introduction

Shadowing has become a well-known and recognized technique for acquiring native speakers' pronunciation and intonation in foreign languages [3], although demands teachers' supervision. This research aims to study a method to establish rules for development of an automatic feedback system for language learners.

The objectives of the research are the following:

(1) To find out mistakes, that are common among the learners of Japanese and significant to native speakers.

(2) To find out acoustic speech characteristics of the shadowed speech to discover the rules for the future feedback system.

2 Related Works

The need of visual feedback on pronunciation for L2 learners has been discussed in various works, dedicated to Spanish, English, and other languages. There are some research works on visual feedback criteria [2], but not related to specifics of the Japanese language. Presently available tools for shadowing have no feedback option, except only showing a waveform of the utterances.

3 Experiment

3.1 Speech Material

ATR503 sentences speech database was used as the source of Japanese speech material. We chose recordings of 10 sentences as the "reference utterances" for shadowing exercises: read by one male and one female native speaker. Each of the set consisted of one and the same list of utterances, pronounced at normal speed (natural reading) and at slow speed (phrase-segmented reading).

Six foreigners living in Japan for 1-3 years took part in the experiment: 3 men and 3 women. Their levels of Japanese language proficiency ranged from JLPT N5 to JLPT N2.

AUDACITY was used for the learners' shadowing recordings on the R-channel of the speech data while reference utterance was on the L-channel, in order to keep the relative timing between reference and learner's utterance. JULIUS was used for phoneme segmentation and PRAAT served us to obtain fundamental frequency contour of the utterances.

3.2 Evaluation by Native Speakers

Two types of questionnaire forms allowed to receive qualitative and quantitative evaluation data from 15 Japanese native speakers, 2 of them were university media club members who practiced public speaking skills. In Questionnaire 1, they evaluated mistakes in pronunciation of sounds in the words, in pitch accent in intonation pattern and commented on anything unnatural in the whole utterance or in a part of it. Questionnaire 2 was focused only on intonation in words with particles before small pauses and in the end of utterances.

4 Results

4.1 General Findings

A comparative analysis of the Questionnaire 1 answers with the acoustic features of the recorded utterances revealed common mistakes in the recorded speech of learners. They can be divided into two groups, related to the type of utterances:

1. Mistakes in pharse-segmented reading:
   - flatness of the intonation of the whole utterance
   - mistakes in word accent (pitch)
   - mistakes in pitch at particles
2. Mistakes in natural reading:
   - mistakes in word accent (pitch)
   - mistakes in mora duration

There were less commentaries on the intonation patterns in pronouncing at normal speed compared to the slow speed utterances. However, the omission and mispronouncing of words to keep the reference pace were common as expected, which complicates correct labeling of utterances and hence detecting mistakes automatically.

Mora duration can be compared via dynamic time warping (DTW) analysis. As pitch accent depends on intonation in Japanese [2], their correctness can be analyzed based on calculation of fundamental frequency (F0) contours.

We see the flatness of the intonation to be one of the easiest characteristics to investigate. We can measure the ratio of the intonation changes (octave/s) that are particular to a learner and can comment on the rising or lowering the intonation before pauses within utterance or in its end.

4.2 Mistakes in Intonation

A comparative analysis of the automatic calculations of F0 characteristics and answers to the Questionnaire 2 allowed to define a rough range within the intonation deviations which are not considered by native speakers to be as significant mistakes. The possible ratio for phrase-segmented reading and for natural reading are from - 3.0 to 2.0 and from - 2.0 to 2.5, respectively (Fig. 2 and 3).

5 Conclusion

In this study, we proposed a method for detection and correction of the mistakes in recorded speech of learners of Japanese based on the analysis of fundamental frequency numbers. The efficiency of the method was verified and modified with the reference to the native speakers' evaluation.

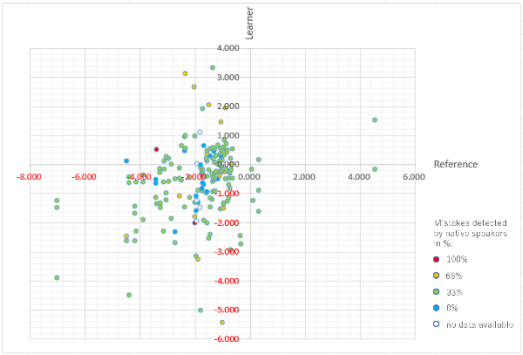

Fig. 2 Distribution of learners' mistakes in intonation in phrase-segmented reading

Fig. 3 Distribution of learners' mistakes in intonation in natural reading

References

[1] Daniel Olson, J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. Language Learning & Technology, 18(3), 173–192. Retrieved from http://llt.msu.edu/issues/october2014/olson.pdf

[2] Isamu Abe, "Intonation in Japanese," in Intonation Systems - A Survey of Twenty Languages (eds. by D. Hirst & A. di Cristo). Cambridge University Press, 1998.

[3] Lambert Sylvie, "Shadowing," in Journal des Traducteurs (Translators' Journal), Les Presses de l'Université de Montréal, 37(2), pp. 263– 273, 1992.