Development and Maintenance of Practical and In-service Systems for Recording Shadowing Utterances and Their Assessment

Junwei YUE†, Daisuke SAITO†, Nobuaki MINEMATSU†, Yutaka YAMAUCHI‡

†The University of Tokyo,   ‡Tokyo International University

{jwyue,dsk_saito,mine}@gavo.t.u-tokyo.ac.jp, yyama@tiu.ac.jp

Abstract

We demonstrate two systems, network-based and standalone, for collecting shadowing utterances and their automatic assessment. Since June 2016, these systems have been used in real English classes at several universities in Japan. The following features are highlighted: 1) To avoid pop noises from a speaker and reduce babble noises from other surrounding students, an ear-hook microphone is used with a USB audio device, 2) A network-based system and a standalone system were developed separately because network traffic not rarely causes technical errors when recording, 3) For beginners, easy-to-understand illustrations are prepared for them to get accustomed rapidly to shadowing practices, 4) DNN-based GOP calculation is run and its score is fed back to learners, and 5) To motivate them, the GOP score distribution over the learners are also fed back for them to compare their own scores with others'. In demonstration, each feature will be exhibited and explained in detail.

Index Terms: Shadowing, recording, assessment, ear-hook microphone, DNN, GOP, feedback instructions

1. Background

For good and valid assessment, an adequate recording environment and an encouraging learning environment are essential. We developed two systems, network-based and standalone, for online and offline learners with several features enabling stable recording and motivating learners to practice shadowing repeatedly. These systems are now used in universities in Japan.

2. Network-based system

It allows learners to record shadowing speeches online by themselves. This system is developed using *PHP* and *Javascript* and recording is possible in class and at home. A screenshot of this system is shown in Fig. 1. In actual use of this system, a web page of instructions for recording is prepared so that learners can become skillful in recording. A unidirectional earhook microphone is used to suppress babble noise. To hear model utterances, canalphones are recommended to prevent the model utterances from being leaked and recorded (Fig. 2). Each recorded sentence is converted into wave format and sent to the laboratory server automatically so there is no need to collect them again, which is convenient for teachers and engineers. As for automatic assessment, it will be described in Section 4.

3. Standalone system

Network connection is not always available. For such learners, another standalone system was developed, which has more customizability than the network-based system. In our network-based system, customization often requires operations of a system manager, but in the standalone system, a user, i.e., teacher, can customize the system in an easier way. When our network-based system was tested in some universities, due to network

Figure 1: *A screenshot for the network-based recording system. The button inside red frame is for GOP calculation.*

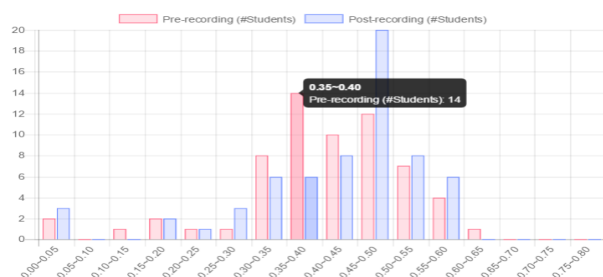Figure 2: *Recording devices used in recording.*

Figure 3: *An example of the histogram of GOP scores in a class.*

traffic, a recording process was sometimes stopped unexpectedly, which discouraged learners. For secure recording, the standalone system is recommended.

4. Online GOP calculation

To motivate learners, the network-based system is capable of calculating the DNN-based GOP score for each shadowing utterance (Fig. 1). The GOP score is calculated using DNN-based acoustic models, which are trained using WSJ corpus [1]. After recording any sentence, learners can know his/her score by a single click. To motivate learners in a class, a tentative web was built to let them know the GOP score distribution of all the learners. The location of each learner was indicated in the distribution (Fig. 3). Every learner can know how well he/she performs shadowing compared to his/her classmates.

5. References

[1] J. Yue, F. Shiozawa, S. Toyama, Y. Yamauchi, K. Ito, D. Saito, N. Minematsu, "Automatic scoring of shadowing speech based on DNN posteriors and their DTW," in *INTERSPEECH*, 2017. (accepted)