GenieTutor: A Computer Assisted Second-Language Learning System based on Spoken Language Understanding

Oh-Woog Kwon, Kiyoung Lee, Yoon-Hyung Roh, Jin-Xia Huang, Sung-Kwon Choi, Young-Kil Kim, Hyung Bae Jeon, Yoo Rhee Oh, Yun-Kyung Lee, Byung Ok Kang, Euisok Chung, Jeon Gue Park, Yunkeun Lee

Automatic Speech Translation and Artificial Intelligence Research Center, Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Yuseong-gu, Deajeon, 305-700, Korea

1 Overview

This demonstration paper introduces GenieTutor - a computer assisted second-language learning system using spoken dialog processing technology. GenieTutor plays the role of a language (English at present) tutor by automatically correcting grammar and checking content properness of the learners' responses, and giving educational feedbacks to learners. The speech recognition system is optimized for non-natives as well as natives for educational purpose and smooth interaction.

GenieTutor leads dialogs with learners to focus on the certain topics by asking questions and providing suggestions for the answers. The system recognizes the speech which is answered in second-language, evaluates if it is the proper answer for given question, checks grammatical errors, and provides feedbacks to help learners practice their English proficiency. Fig.1 shows the schematic diagram of GenieTutor.
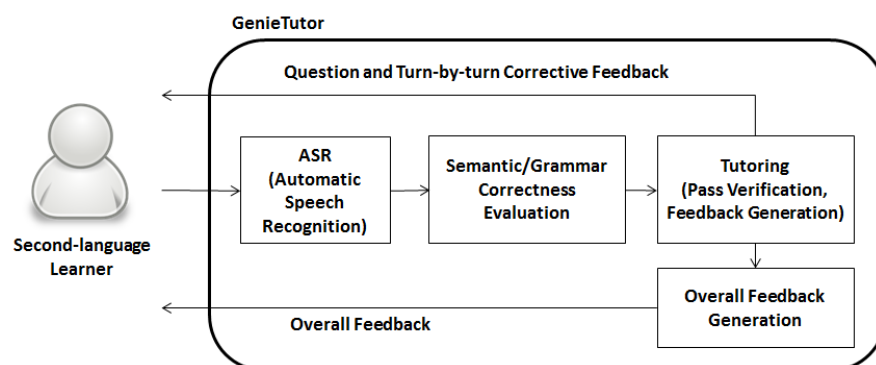

Fig. 1 Schematic diagram of GenieTutor

Non-native-optimized Speech Recognition

In order to construct non-native-optimized speech recognition system, we used native English utterances (380 hours in total) and non-native Korean spoken English utterances (408 hours in total) where each utterance was sampled at a rate of 16 kHz. We first extracted speech feature vectors at every 10 ms for a 20 ms analysis window [1, 3]. Then, we trained separately native acoustic models (AMs) and non-native AMs by using native and non-native utterances respectively [4]. The AMs are composed of 3-state, 16-mixture, and cross-word triphone models. Next, native and non-native AMs were merged based on Gaussian mixture models based on multi-space probability distribution.

In addition, we collected about 130 million English text sentences including the English scripts of GenieTutor, the spontaneous sentences, and the grammatically wrong sentences that are commonly occurred by Korean learners. Then, we constructed a back-off trigram language model with 54,826 most frequent words.

Evaluation and Tutoring

Evaluation module evaluates whether the learner's utterance is appropriate semantically and is correct grammatically. The semantic correctness checker decides whether to pass the learner on the current turn or not, using the domain knowledge and language model. The semantic correctness is classified into 6 categories such as "perfect", "too few modifiers", "inflection error", "subject-verb error", "content error", and "illegal expression", for the feedback to the learner in the tutoring step.

Grammatical error correction plays an important role in second language learning using computers. Many grammatical error correction systems aim at detecting and correcting grammar errors in essays written by students who are non-native speakers of English [2]. We focus on grammar errors in dialogue between a student and a learning system. For a grammatical evaluation, we employ a grammatical error correction system which is composed of three different approaches: a rule-based, a machine learning based and an n-gram based correction modules. According to the performance of each correction module for grammatical error types, we assigned a proper weight to correction candidates by each correction module. To suppress false alarms which are critical to secondary language tutor system, the correction system has a voting strategy to filtering implausible correction candidates.

Tutoring module decides to go to the next turn or ask the learner to try again according to the semantic correctness evaluation, and also provides a turn-by-turn corrective feedback using evaluation results. The feedback consists of three parts that are shown to the learner in a step-by-step and sequential manner. The first part briefly shows a grade of pass or fail and the words with grammar errors. The

second part suggests some recommendation sentences with reasons of failure when the semantics of the input utterance is not appropriate to the current question. The last part is the corrective feedback of grammatical errors described in the first part.

Once the dialogues between the GenieTutor and the learner finished, overall evaluation module assesses the English learner's performance, and produces feedbacks with the scores to show which part the learner should focus more on. Several measurements are adopted for the evaluation, including task proficiency, grammar accuracy, vocabulary diversity, and grammar complexity. Task proficiency evaluates how fluently the conversation has been maintained according to the numbers of pass and failing turns. Grammar accuracy is scored based on the number of grammar errors, which is checked by grammar check module. Grammar complexity compares user's sentences with the references offered by native speakers, and gives relative complexity score by considering the length and the number of conjunctions. For the vocabulary diversity the system is supposed to check if the learners tend to use the same expressions or words from elementary level vocabularies when there are better alternatives, and provides synonyms and similar expressions to improve learners' vocabulary.

2. Curriculum of GenieTutor

The service or curriculum of GenieTutor consists of three stages such as Pronunciation Clinic, Think&Talk and Look&Talk.

Pronunciation Clinic

The Pronunciation Clinic provides the English pronunciation skills that are specialized for Korean non-native language learners and the corresponding speaking practices based on speech recognition. As shown in Fig. 2, the Pronunciation Clinic consists of four steps; (a) selecting a lesson among the 30 Pronunciation Clinic lessons, (b) listening a video lecture for the selected lesson, (c) selecting the words or sentences that contains the target pronunciations of the selected lesson, and (d) practicing a spoken dialogue by using the selected words or sentences. Especially, the last speaking practice step provides pronunciation scores and the comparative analysis results of the intonation and stresses between the learners' and the corresponding native utterance.
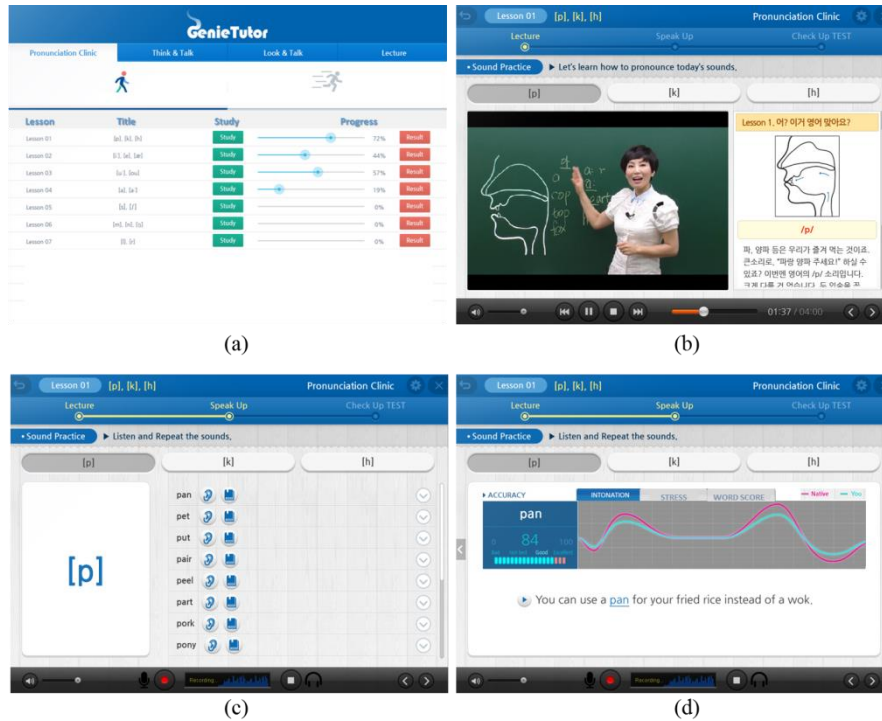
(a)

(b)

(c)

(d)

Fig. 2. Four steps of Pronunciation Clinic

Think&Talk: Talk with the computer on various subjects

In the stage Think&Talk an English learner is supposed to select a dialog context from a pool of domains and its contents, and to practice the dialog utterance that corresponds to the context. If the learner makes a semantic or grammatical mistake, GenieTutor gives him some corresponding feedbacks and asks him to try again a correct utterance.
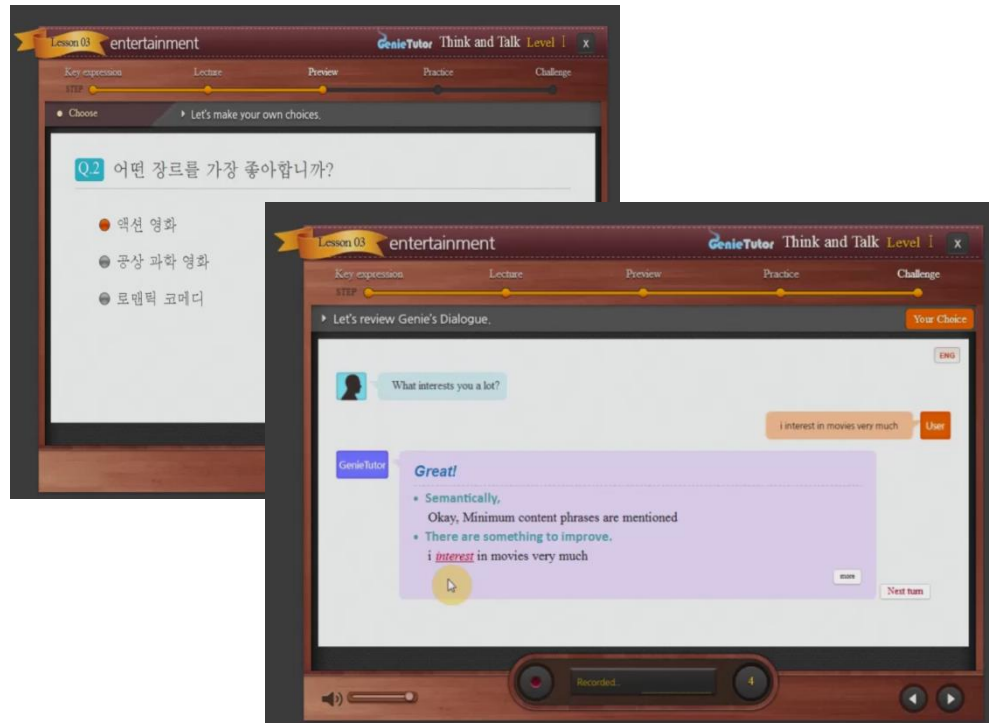
Fig. 3. Instance selection and a dialogue exercise in the stage Think&Talk

Look&Talk: Look and describe the pictures to the computer

In the stage Look&Talk an English learner is supposed to watch and describe the picture according to the system's question to enhance his/her power of expression. Likewise with the stage Think&Talk, GenieTutor gives feedbacks to the learner and asks him to try again a correct utterance if the learner makes a semantic or grammatical mistake.
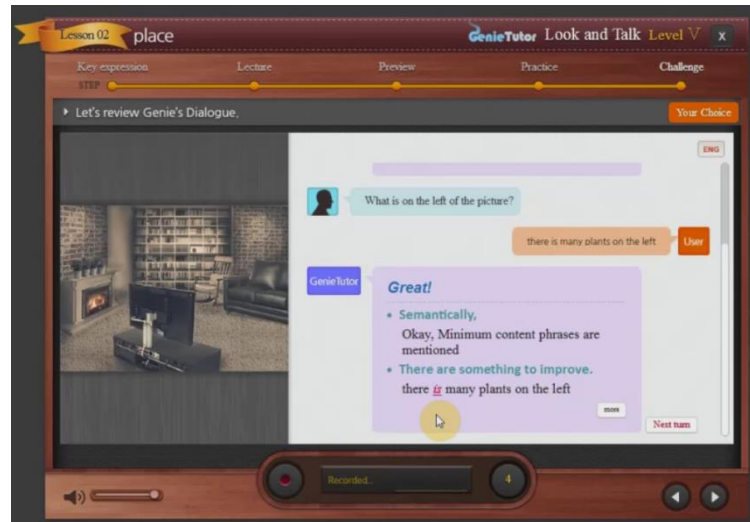
Fig. 4. A dialogue exercise of the stage Look&Talk

Reference

1. Hoon Chung, Sung Joo Lee, Yun Keun Lee, Weighted finite state transducer-based endpoint detection using probabilistic decision logic. ETRI J 36:714-720. (2014)
2. Hwee Tou Ng, Siew Mei Wu and Ted Briscoe, The CoNLL-2014 Shared Task on Grammatical Error Correction. In *Proceedings of the 18th Conference on Computational Natural Language Learning* , pages 1-14. (2014)
3. Sung Joo Lee, Byung Ok Kang, Hoon Chung, Yun Keun Lee, Intra- and Inter-frame features for automatic speech recognition. ETRI J 36:514-517. (2014)
4. S. Young, et al., The HTK Book (for HTK Version 3.4). Microsoft Corporation, Cambridge University Engineering Department, Mar. (2009)