A CALL system for teaching the duration and phone quality of Japanese tokushuhaku

Goh Kawai and Keikichi Hirose

Department of Information and Communication Engineering, University of Tokyo
7—3—1 Hongo, Bunkyo—ku, Tokyo 113—8656, Japan

goh@kawai.com  hirose@gavo.t.u—tokyo.ac.jp  http://www.gavo.t.u—tokyo.ac.jp/

Abstract: A CALL (computer—aided language learning) system was developed to teach the pronunciation of Japanese tokushuhaku (long vowels, the mora nasal and mora obstruents) to beginning—level learners whose native language is American English. Nonnatives often produce tokushuhaku with incorrect phone duration or quality. The proposed system (a) detects mistakes in phone duration and quality, (b) returns an intelligibility score that tells the learner the likelihood of native speakers understanding the learner’s pronunciation the way it was intended, and (c) suggests ways to improve pronunciation.

INTRODUCTION

Acquiring the pronunciation of individual phones is a crucial requirement in nonnative language training because the pronunciation of entire utterances depends heavily upon the correct pronunciation of each phone. Unfortunately current methods of teaching are inefficient because they rely on human instructors who in many cases cannot spend much time with each student. Using CALL to teach many students individually in parallel may increase skill acquisition and retention by shortening the time required to detect and correct pronunciation errors early in the students’ learning careers.

The predominant phonemic contrast between Japanese tokushuhaku is phone duration. As few non—Japanese languages distinguish phonemes on the basis of duration, acquiring tokushuhaku is a major challenge in TJSL (teaching Japanese as a second language). This paper proposes a method to detect errors in phone duration and quality of tokushuhaku spoken by native speakers of American English learning Japanese as a second language. Our system adopted American English as L1 and Japanese as L2, but our techniques can be applied to other language pairs.

SYSTEM OVERVIEW

Figure 1 shows the system–user interaction. Figure 2 shows the process flow of the system. The learner is prompted by the system to read tokushuhaku minimal pairs aloud. Tokushuhaku phones are identified using speech recognition in forced—alignment mode (i.e., phone boundaries are determined given a correct transcription of the utterance or similarly tightly—constrained language model). Two separate measurements are made: the first measures the intelligibility of tokushuhaku based on their duration, and the second sees whether non—Japanese phones were inserted or substituted in the utterance.

Phone durations are compared with results from perception experiments that measured the intelligibility of tokushuhaku with respect to changes in phone duration [1]. The learner obtains a score that tells him the percentage of native speakers who will understand his tokushuhaku based on its duration.

FIGURE 1. System—user interaction

FIGURE 2. System process flow

2981
Mistakes in phone quality are detected using bilingual context-independent phone models of the learner's L1 (native language — American English [2]) and L2 (target language — Japanese [3]). HMMs for L1 and L2 are trained separately on language-dependent speech data, but during recognition, L2 phones and L1 phones that are often erroneously substituted with L2 phones are grouped together into confusion sets. These meta-phones are designed according to pedagogical knowledge. For instance, to capture a common mistake of diphthongizing or reducing vowels, we created meta-phones such as:

\[
\text{meta}_a = \{ j_a, j_a:, e_aa, e_ae, e_ah, e_aw, e_ay, e_ax, e_axr \}
\]

\[
\text{meta}_o = \{ j_o; j_o: ; e_ao; e_ow; e_oy ; e_ax, e_axr \}
\]

Similar meta-phones where made for other phones. Forced-alignment was performed by using meta-phones in place of phones. The closest phone recognized indicates the nonnativeness of the utterance should the recognized phone not be Japanese. Prior knowledge of L1 and L2 can identify nonnative articulatory gestures that result in pronunciation errors, thus allowing us to estimate the learner's manner of articulation and to suggest a remedy. An example of corrective feedback is included in Figure 1.

SYSTEM PERFORMANCE

The system detects mispronounced tokushuhaku duration and quality, such as shortened or lengthened tokushuhaku, diphthongized long vowels, palatalized plosives and inserted liquids. Figure 3 shows an example of forced-alignment results of the word “kado” pronounced in three ways. Phone labels prefixed with “j_” denote Japanese phones, and “e_” denote American English phones. Recognition results closely match human judgements. Figure 4 shows how a subject’s intelligibility score improved over time as he practiced tokushuhaku duration for the word “kado”.

CONCLUSION

The paper discussed a CALL system for the teaching of tokushuhaku phone duration and quality to native speakers of American English. By using phone models for other languages, teaching phone duration and quality of other languages may be possible. Future work is planned on teaching fundamental frequency contours, such as lexical pitch accent and pitch contours of set phrases.

ACKNOWLEDGMENTS

We thank Steve Young for providing us with American English HMMs [2]. We appreciate Kazuya Takeda’s providing us with Japanese HMMs [3].

REFERENCES


2982