

7. CONCLUSION

This dissertation advances the state of the art by combining existing technologies with newly created knowledge. The research tasks provide multiple viewpoints towards the larger issue of automated pronunciation learning. CALL systems were built to teach three pronunciation skills: phone duration, pitch, and phone quality. This section lists our major findings and directions for future work.

7.1. FINDINGS

The system for teaching phone duration provides corrective feedback similar to human teachers by measuring phone duration using speech recognition technology. We learned that (1) learners adjust their speaking rate to the system's prompts, (2) native speakers distinguish long and short tokushuhaku clearly, (3) instructing learners to say a phone longer or shorter is easy to understand, and (4) learners rapidly acquire skills.

The system for teaching phone quality provides corrective feedback while disregarding individual physiological aspects of the learner's speech by measuring L1 accents by comparing L1 and L2 phones using a speaker-independent bilingual phone recognizer. We learned that (1) the number of phone types in the learner's interlanguage allophone inventory decreases with pronunciation practice, (2) the size of the learner's interlanguage allophone inventory is one measure of the learner's level of foreign accent, (3) the frequency distribution of interlanguage allophones being selected as the L2 phone is another measure of the learner's level of foreign accent, and (4) the system helps learners rapidly acquire phone quality skills.

The system for teaching pitch contours automatically grades prosody by measuring pitch patterns and intonation contours using speech recognition technology and prosodic analysis. We learned that (1) native speakers clearly distinguish pitch changes at syllable boundaries, (2) instructing learners to say a syllable higher or lower is easy to understand, and (4) learners rapidly acquire skills.

7.2. FUTURE WORK

Teaching phone duration can be improved by gradually lowering intelligibility scores as the durations of long phonemes become overly long. This study ignored this aspect because we found that learners do not elongate long phones to the extent of sounding unnatural or incorrect (undoubtedly because learners wish to comply with target language norms). However at least in theory such erratic behavior is possible. Perception experiments using synthesized long phones can determine the goodness of phones.

A related issue involves multimorpheme superlong tokushuhaku, such as the 6-mora-long [o] in “denwa bangoo o oo ojini osieru” (give the telephone number to my great uncle). Superlong tokushuhaku occur across multiple morphemes. Morpheme boundaries are perceived by pitch accent patterns. Superimposing pitch accent on phone duration instruction may help learners acquire skills in this area.

The pitch contour system could be improved by teaching pitch contours of phrases and sentences with more than one acceptable intonation contour. Here again the concept of quantitative measurements of intelligibility rears its head. Manageable tasks might include categorizing various intonation contours into various meanings — for instance “kawai san desuka” with a falling tone means “So you are Mr. Kawai,” (indicates confirmation of information) while a rising tone means “Are you sure you are Mr. Kawai?” (indicates incredibility). Simple tasks such as these can be based on perception experiments. More advanced tasks might include intonation as discourse structure markers — for example utterance-final intonation patterns functioning as cues for turn-taking behavior.

Teaching phone quality might be significantly improved by quantitatively measuring the intelligibility of phones as a function of phone quality. On the one hand, we could perhaps extrapolate intelligibility from rank-order or qualitative judgements of native speakers. On the other hand the correlation among native speakers regarding pronunciation has never been high; in fact this has been the limiting factor for many pronunciation grading systems. Thus unfortunately the goodness of phone quality may never be quantifiable.

A fundamental issue that this dissertation has not touched upon is the training of hearing skills. Spoken language competence necessarily requires both production and reception skills. This dissertation has not dealt with teaching speech perception to non-native learners primarily because hearing skills have been the focus of CALL research up to this point; this paper attempts to balance the scale of research efforts by proposing learning systems for speech production. Certainly this paper does not intend to minimize the importance of speech perception training. In fact the systems proposed in this paper would be meaningless without matching teaching in listening comprehension. An example of this would be tokushuhaku training, where learners must learn how to produce long and short vowel lengths correctly, but must also learn how to tell them apart in connected native speech because where localized fluctuations in speech rate can obliterate clear durational distinctions.