CHAPTER 1

INTRODUCTION

1.1 Background

Computer Assisted Language Learning (CALL) System becomes popular tools to learn pronunciation in the second language (L2) because they offer extra learning time and material as well as the possibility to practice in a stress-free environment. With the integration of Automatic Speech Recognition (ASR) technology, these systems, which we will refer to as CAPT (Computer Assisted Pronunciation Training) Systems, can even provide limited interaction: the computer understand the student's speech and react accordingly, thus making the learning process more realistic and engaging, and can provide the feedback on the quality of the student's speech. While students generally enjoy learning with speech-enabled systems, a number of researchers and educators are skeptical about the usability of ASR for pronunciation training in the L2 learners, because this technology still suffers from a number of limitations.

For this reason, several attempts have been made to establish the effectiveness of ASR technology for CAPT. In many publications that

have appeared in the language learning community, criticism has been expressed with regard to the two main features of ASR that is to be used by language learners: the ability to recognize accented or mispronounced speech, and the ability to provide meaningful evaluation of pronunciation quality. In the following chapters we propose some methodology to detect the pronunciation errors of the student's speech and design a number of experiments to examine the usability and performance of the CAPT system.

1.2 Automatic Speech Recognition for CAPT

The ideal ASR-based CAPT system can be described as a sequence of five phases, the first four of which strictly concern ASR components that are not visible to the user, while the fifth has to do with broader design and graphical user interface issues. 1) Speech recognition: the ASR engine translates the incoming speech signal into a sequence of words on the basis of internal phonetic and syntactic models. This is the first and most important phase, as the subsequent phases depend on the accuracy of this one. Besides, this phase alone already allows devising a range of computer-based activities to train communicative skills in the L2, such as interactive dialogues with the computer and speech-enabled multiple-choice exercises. However, the main pedagogical advantage that ASR-based CAPT can offer for training oral skills in the L2 is the provision of an evaluation of pronunciation quality. The following phases show how this evaluation is possible. 2) Scoring: this phase makes it possible to provide a first, global evaluation of pronunciation quality in the form of a score. The CAPT system analyses

the spoken utterance that has been previously recognized. The analysis can be done on the basis of a comparison between temporal properties (e.g. rate of speech) and/or acoustic properties of the student's utterance on one side, and natives' reference properties on the other side: the closer the student's utterance comes to the native models used as reference, the higher the score will be. The usefulness of automatic scoring for pronunciation training is evident, as it gives the learner immediate information on overall output quality and on how this can improve over successive attempts. 3) Error detection: the system can locate the errors in the utterance and indicate to the learner where s/he made mistakes. This is generally done on the basis of so-called confidence scores that represent the degree of certainty of the ASR system that the recognized individual phones within an utterance actually match the stored native models used as a reference. Signaling that a certain sound within a word is problematic can be particularly useful to raise awareness in the learner of that problem and thus help her/him to focus and practice more on that area. 4) Error diagnosis: the ASR system identifies the specific *type* of error that was made by the student and suggests how to improve it, because a learner may not be able to identify the exact nature of his pronunciation problem alone. This can be done by resorting to previously stored models of typical errors that are made by non-native speakers. 5) Feedback presentation: this phase consists in presenting the information obtained during phases 2,3, and 4 to the student. It should be clear that while this phase implies manipulating the various calculations made by the ASR system, the decisions that have to be taken here – e.g. presenting the

overall score as a graded bar, or as a number on a given scale – have to do with design, rather than with the technological implementation of the ASR system. This phase is fundamental because the learner will only be able to benefit from all the information obtained by means of ASR if this is presented in a meaningful way.

1.3 Research Topic

In the CAPT system, it is well known that vowel pronunciation is much more important than that of consonants. Successful CAPT applications have been established, but few of them focused on the vowel pronunciation learning, especially for the text-independent pronunciation learning (which does not require the use of a target sentence).

Our goal is to construct a system to find the vowel error pronunciation in the utterance of L2 learners and generate instruction of English vowel pronunciation for Taiwanese learners. In this thesis, we propose a pronunciation assessment method based on HMM and formant coefficients, which is able to give high-level instructions and assessments about the articulator. To do so, we design a number of experiments to verify the performance of the methods we proposed.

There are three main part of our system, 1) speech segmentation, it refers to divide speech waveform into phonemes. 2) formant-based hmm training, in addition to 39-dimentions MFCC, we also add formant coefficients as a new feature set for HMM training .3) formant-level assessment, a more robust pronunciation error detection methodology is present in this research. Figure 1-1 shows the system framework.

First, we use pronunciation confusion network (PCN) to predict the pronunciation error of L2 learners. Then the normalized formant frequency of each phoneme can be calculated according to each student's maximum and minimum formant frequency. The goal of formant-level assessment is to compute the confidence score of each phoneme from incoming utterance, and we compare the score with threshold to determine if the phoneme is pronounced correctly.

1.4 Related Work

Recently Yasushi et al. has proposed a formant-based CAPT system for vowel learning [16]. They described a knowledge-based approach to generate instructions according to the formants of vowels. The mapping of the formants from a native speaker to a second language (L2) learner is built and an ideal formant is expected according to the preliminary process of the L2 learner's native language. A proper instruction is then responded according to the difference between the ideal and the actual formant frequencies.

However, the formant is speaker dependent and a global mapping is not always possible, particularly for the formant of native speaker. Besides, the co-articulation between consonant and vowel may also make the formant look different. Therefore, in the formant-level assessment, the co-articulation between phonemes and the speaker dependent formant normalization are essential. The proposed system takes these factors into consideration and tries to create a robust computer-assisted learning system for vowel pronunciation, for Chinese in Taiwan. Related research on spoken English learning for

Chinese speaking people is seldom reported in the literature.

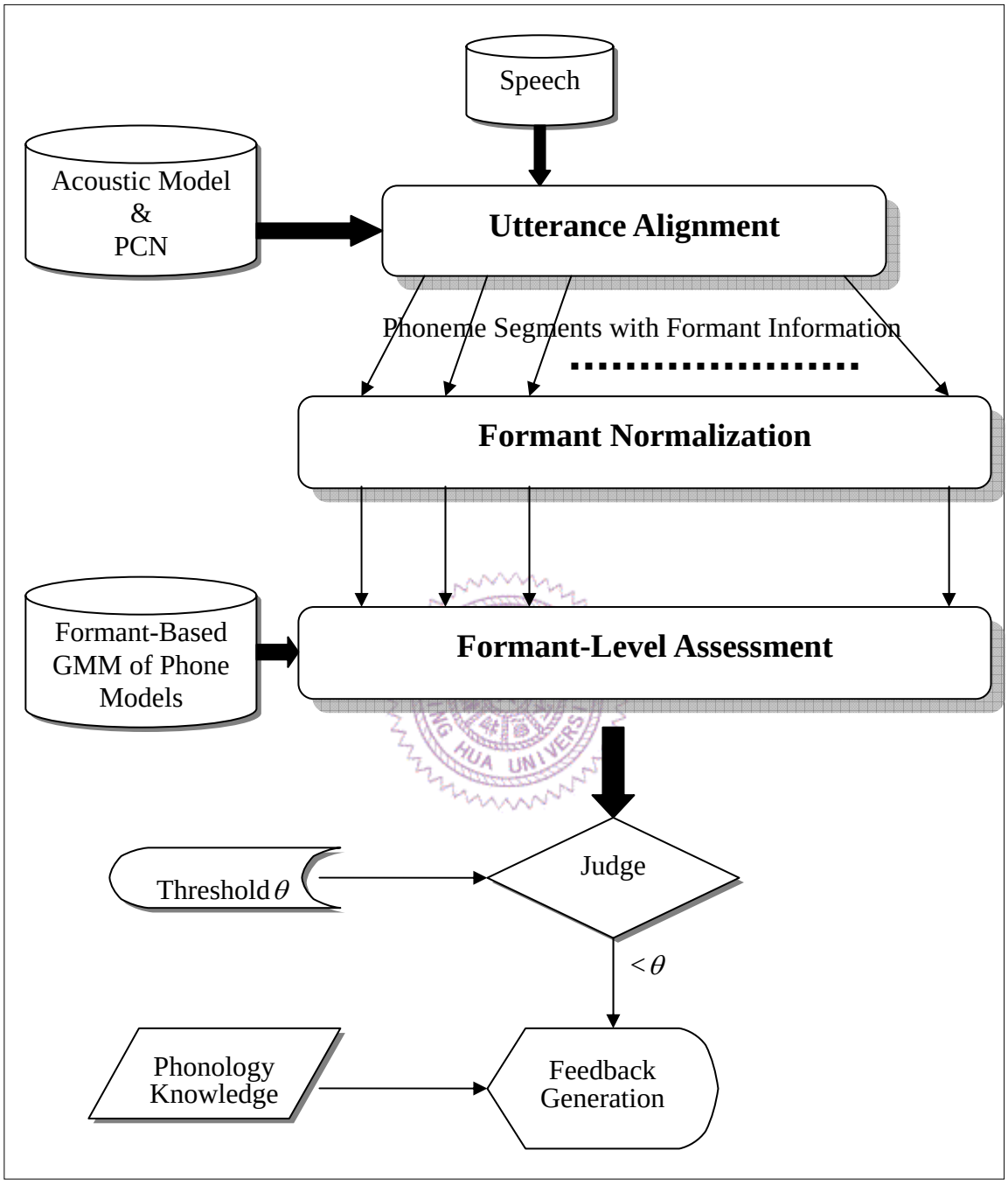

Figure 1.1 System Framework