THE FEATURES OF CHINESE COMPUTER-AIDED LANGUAGE LEARNING SYSTEM

YUE Dongjian, CHAI Peiqi
AI Laboratory, Department of Computer Science and Technology, Tongji University, Shanghai
yuedjk@online.sh.cn  http://cs.tongji.edu.cn/ai

ABSTRACT
In this paper, we firstly analyse the trend of spoken language or speech learning at present. We then investigate the problems of applying speech technology in language learning and key techniques to be used. According to the features of Chinese spoken language, we propose the rules and methods when a Computer Assisted Language Learning (CALL) system for learning Chinese is designed or realized.

1. INTRODUCTION
Speech technology (such as speech recognition, synthesis, analysis, manipulation and visualisation) offers great promise in many aspects of language learning by human. It has already been used in application such as pronunciation training for hard-of-hearing persons and is now appearing in second language learning. As more advanced speech technology emerges, its usefulness will increase. Speech Technology in Language Learning (STiLL) is an important trend in spoken language processing area currently.

Many researches show that: Chinese is quite different from many western languages in various structural features. The monosyllabic structure with tones, the open vocabulary nature, the flexible wording structure, and the flexibilities in word ordering are good examples of the structural features of Chinese language. It is believed that better results and performance will be obtainable in developing spoken Chinese language processing technology, if such structural features of the language can be well considered.

This paper summarizes the most aspect of speech technology in language learning firstly, taking account of the structural features of Chinese language, presents an integrated framework of CALL systems for Chinese spoken language learning.

2. THE MAIN CONTENT AND TECHNIQUES IN CALL
The main content and key techniques of applying speech technology in speech learning is proposed and summarized here. The relation of them in a CALL may be described in Figure 1.

2.1 Two Categories of Pronunciation Training:
Pronunciation training of any spoken language can be divided into two categories. One is the training of speech unit, such as phoneme, vowel, consonant, syllable, word, phrase and sentence etc. Another is apperceiving and training of prosody, such as accent, tone, intonation, tune, dialect.

2.2 Available Relationship Between Scorings:
The pronunciation scoring of words or phrases is used to determine how well they were said. Many speech recognition system have a score as a "by-product". However, this score is tuned for used by native speakers. It may does not work well for language learner. The reason is that the models are typically trained on non-native speech and on a homogeneous dialect. So validating such scores to improve learner’s pronunciation is very important. The way is to establish a correct corresponding relationship between the scoring which teachers or experts gives to students and the scoring which programme creates for learner.

2.3 Scoring Calculation
There are three main kinds of speech recognition technology, which can be used in the calculation of scoring. One fashion method is the HMM-based speech recognition technique. Such as the HTK (HMM) is successful software, which is developed by Entropic Lab. of Cambridge University. The HMM block used in IBM 'via Voice' is also this kind. The technique of speech spectrum comparing is used in IBM 'Speech View'. The third method is comparing of formants.

For examples, log-likelihood scores and log-posterior probability scores are two kinds of phonetic scores which may be calculated in HMM[8].

- Log-likelihood scores. For each phone segment \( i \) (of \( d \) frames: \( [t_0, t_0+d] \)), the log-likelihood score is defined as \( l \) where \( p(y,t|q) \) is the probability that the observation vector at time \( t \) is generated via the phonetic model \( q_i \):

\[
l = \frac{1}{d} \sum_{t=t_0}^{t_0+d-1} \log p(y,t|q_i) \quad (1)
\]

- Log-posterior probability scores. The scores of a phone segment \( i \) is computed as the average (over time) of the frame based posterior probability of the phone \( q_i \) at time \( t \). \( N \) stands for the number of phonetic models and \( P(q,t) \) is the prior probability of the phone class \( q \):

\[
p = \frac{1}{d} \sum_{t=t_0}^{t_0+d-1} \log \frac{p(y,t|q_i)P(q_i)}{\sum_{j=1}^{N} p(y,t|q_j)P(q_j)} \quad (2)
\]
2.4 Prosody Verification

Experience shows that a person with perfect phone pronunciation who lacks correct timing and pitching is very hard to understand. The “song” that the speaker “sings” while emitting a string of phones is the “glue” that holds the whole message together, guides the listener along, indicates where important content words are, disambiguates parts of sentences, and enhance the meaning with style and emotion.[2].

The main content of verifying the correctness of prosody includes such work as calculating of pitch or fundamental frequency contour, duration (speech signal length) and intensity. If the speech is phrase or sentence, the HMM-based algorithm must be used firstly to realize the segmentation and transcription of speech.

2.5 The Visual Feedback

The pronunciation quality can be feedback to instructor and learner with a serial of graphs, such as still Vowel Chart, dynamic Vowel Chart and F1-F2 Chart. These charts are effective for learner to improve their pronunciation.

There are several rules can approximately depict the relationship among formants and the features of vocal tract during pronunciation.

- The frequency of F1 is inversely related to tongue height (high vowels have a low F1).
- The frequency of F2 is related to tongue advancement (F2 increases as the tongue moves forward).
- The effect of lip rounding is to generally lower the formant frequencies of F1, F2, F3, F4.
- The fourth formant, F4, is nearly constant in frequency for any one speaker.
- The formant bandwidths according to House and Stevens (1950) are: F1: 54Hz, F2: 65Hz, F3: 70Hz.

2.6 Prosody Perception and Modification

In learning and apperceiving the prosody of speech, there are about three kinds of models used as reference: a speech database, a phrase recorded by the teacher in the classroom and text-to-speech synthesis system.[7] The two main prosody perception and modification tools are the pitch contour and resynthesized speech with TD-PSOLA may be used as feedback information.

2.7 Others

- There are mostly two form or style in designing STiLL software: Couse and Game.
- Some syllables may be difficult in pronunciation. In order to learning such pronunciation, a set of minimum pair samples can be used by learner to compare and apperceive such syllable.
- Interactive style is effective in speech learning programme.

![The Main Aspects of Speech Technology in CALL](image)

Figure 1. The Relation of Main Aspects and Technology in CALL
3. THE FEATURES OF CHINESE CALL SYSTEM

According to the features of Chinese phonology[4], we proposal that a Chinese CALL system may be implemented and fined according to two sides as follows:

3.1 The Learning of Syllable and other Small Units:

If one begins to learn Chinese as a foreign language, he always start from small speech units: Chinese phonetic (PinYin) alphabet, the initial and final of syllable, the spell of the initial and final as well as tone of syllable. In this case:

• Three kinds of methods (HMM-based speech recognition, comparing of formants or spectrum) discussed in front paragraph (2.3) can be used to the scoring calculating.

• The speech spectrum, vowel chart, F1-F2 formants chart and simulation graph of vocal cave (2.5) can demonstrate pronunciation result and quality of learner’s speech. The comparison of speech between the template and the sample pronounced by learner may be evaluated to get the difference of them. These informations are very useful to discover the problem of mispronunciation and correct it.

• When training the tones of Chinese syllable, a good algorithm should be applied to get the accurate pitch contour(2.4). The pitch contour created by themselves and by teachers can be displayed in a same graphical window, which may help learner to apperceive tones visually. Another method is to resynthesized the learner’s speech with corresponding pitch contour in teacher’s speech using TD-PSOLA technique (2.6).

3.2 The Learning of Phrase and Sentence:

The second stage of Chinese speech learning is to study the pronunciation of multi-syllable word, phrase and sentence. For the reason of co-articulations and other super-segmentation features, the training is complicated in this case. The content of learning here includes two parts: one is the pronunciation of such ‘large’ units; another is the prosody model and its variety.

• In former, speech recognition technique can be used to implement the segmentation, alignment and transcription of the large units. The score of them may be obtained as 3.1. Dynamic vowel chart and spectrumgraph can be used to provide the feedback information.

• Based on segmentation, the pitch contour and length of every segment can be calculated. The pitch contour, topline and baseline of entire learning unit can be obtained. By comparing or resynthesizing with teacher’s patterns or templates, the interactive and effective learning may be carried out.

Figure 2. One example of Chinese learning with SA.

(a)(Top-Left) Waveform (b) (Top-Right)Formants (c) (Bottom-Left)Pitch Contour (d) (Bottom-Right)F1-F2 Chart
4. EXPERIMENTS

Speech Analyzer (SA) is a general speech analysis tool designed by SIL firm. It may assist users who are engaged in speech analysis. It provides a method for adding a time-aligned IPA transcription to the speech waveform using the 1996 version of the International Phonetic Alphabet. It also supports for adding time aligned Phonemic, Orthographic and Gloss texts. The visual features of this programme may be referenced as to design speech learning software.

Within the general speech analyzer, We have tried to implement the Chinese spoken language learning task according to which we have proposed above and get a good result. One learning example of a Chinese is showed in figure 2. Two words or phrases (tong2 ji4 da4 xue2 ji1 xi1) is included in this example. In the sub-graph (a)-top-left corner, the waveform of these phrases is showed. The alignment and transcription may be implemented by expert or teacher. In the sub-graph (b)-top-right corner, the formants and spectrum of the selected segment is drawed. In the sub-graph (c)-bottom-left corner, the pitch contour of the whole segment is showed. In the sub-graph (d)-bottom-right corner, the F1-F2 Chart is given, one dot in which points out the values of F1 and F2 for the selected segment.

5. CONCLUSION

Many studies and experiments show that speech technology can complement and enhance traditional methods of language learning. Although speech recognition and others speech processing software are becoming popular gradually, its appropriate use in language education still presents many challenges. Technology-based language instruction is a multidisciplinary activity and require cross-cultural communication between engineers and language teachers.[1]

To maximize use of the technological potential requires that technology developers provide tools that language pedagogy experts can use. To develop such tools will require that technologists better understand pedagogical issues, and that language pedagogy experts better understand the technology. Linguistic knowledge can be technologically and pedagogically mined in this respect by examining phonetics, phonology, morphophonology and their integration in communication.

Because Chinese is quite different from many western languages in various structural features such as the monosyllabic structure with tones, the open vocabulary nature, the flexible wording structure, and the flexibilities in word ordering. When a Chinese CALL system is devised or implemented, these features must be considered carefully and detailedly. It is not only the methods to validate the speech technology in language learning, but also is the key of such systems to the success.

6. ACKNOWLEDGEMENT

I would like to thank CiTaL FH-Konstanz German, Prof. W.Thomassen and Prof. P.Slowig, for their engagement and enthusiasm in the study of Chinese CALL system. I also would like to thank Mis A.Mauuer and Ms W.Mallow for their help.

7. REFERENCES