Lai, Y.-S., Tsai, H.-H., & Yu, P.-T. (2009). A Multimedia English Learning System Using HMMs to Improve Phonemic Awareness for English Learning. *Educational Technology & Society, 12* (3), 266–281.

A Multimedia English Learning System Using HMMs to Improve Phonemic Awareness for English Learning

Yen-Shou Lai[1], Hung-Hsu Tsai[2] and Pao-Ta Yu[3]

[1]Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan // lys@cs.ccu.edu.tw

[2]Department of Information Management, National Formosa University, Huwei, Yulin, Taiwan // thh@nfu.edu.tw

[3]Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan // csipty@cs.ccu.edu.tw

ABSTRACT

This paper proposes a multimedia English learning (MEL) system, based on Hidden Markov Models (HMMs) and mastery theory strategy, for teaching students with the aim of enhancing their English phonetic awareness and pronunciation. It can analyze phonetic structures, identify and capture pronunciation errors to provide students with targeted advice in pronunciation, intonation, rhythm and volume. In addition, the paper also applies the mastery learning to effectively help students practice pronunciation of English words and sentences. Finally, this paper adopts a quasi-experimental design and lasting for 12 weeks and 120 third-graders, aged 9-10 years, from an elementary school in Yunlin County in Taiwan. These students were recruited and randomly assigned as the experimental group and the control group, respectively. The former used the MEL system, while the latter received the conventional English teaching. Research data were collected through the Phonemic Awareness Test and the English Achievement Test. The results showed that the experimental group with low phonemic awareness performed significantly better than the control group in the English Achievement Test.

Keywords

Applications in Speech recognition, English pronunciation, Hidden Markov models, Mastery learning, Multimedia learning system

Introduction

Phonemic awareness is an important metalinguistic skill which can let students more effectively acquire reading and spelling abilities (Mehta, Foorman, Branum, & Taylor, 2005). While children learn English, an important step is to train them with high phonemic awareness (Leong, Tan, Cheng, & Hau, 2005; Goswami & Bryant, 1990). Carreker (2005) stated that phonemic awareness training helps remediate the problems of poor spelling at any age. Learners, who possess high capability of phonemic awareness, have better capabilities in pronunciation-recognition, reading, and spelling (Treiman & Baron, 1983). How to promote learner’s phonemic awareness during teaching English has become an essential subject in lecture hour. Also, s/he has more opportunities than others to effectively promote her/his phonemic awareness so as to shorten the learning time of reading and spelling. During teaching pronounces in classes, most teachers in Taiwan often concentrate on teaching learners speech skills. Moreover, they neglect learners’ fostering of the recognition capability of phonemic voice. This leads to the fact that the learners cannot have high pronunciation recognition ability. Therefore, the learners are unable to clearly compare their pronunciation differences with correct ones. It raises the problem of inaccurate pronunciation, late speech development, and low letter knowledge (Mann & Foy, 2007).

While learning English in classrooms, teachers often teach students English pronunciation without computer aids. In recent years, due to the growing advancement of information technologies, large amount of multimedia English learning materials have been developed to enhance the learning performance of English pronunciation (Hincks, 2003). The technology of speech analysis has been used for teaching intonational patterns since 1970s (Zinovjeva, 2005). Therefore, speech analysis has been incorporated in much commercial software for English pronunciation. However, the software is still insufficient in offering the feedback to learners for the analysis results of incorrect pronunciations. Thus, the computer assisted language learning (CALL) systems have been successfully developed to improve those limitations such that traditional CALL systems not only perform speech recognizers but also offer the language learning activity and feedback (Wachowicz & Scott, 1999). Moreover, Precoda, Halverson, and Franco (2000) presented a result that the user interface of the CALL system is designed to give pronunciation feedbacks for the learners’ pronunciation ability, and a conclusion that the design of the user interface plays an important role to attract learners’ attention.

Currently, speech synthesizers and digitally manipulated stimuli have been developed in laboratory studies (Neumeyer, Franco, Digalakis, & Weintraub, 2000). Unfortunately, they are not widely utilized in the design of CALL systems. As a result, linguistic experts, including the Second Language (L2) teachers, exclude to take those software products as tools to teach English pronunciation (Zinovjeva, 2005). The reason is that those systems just provide learners with speech synthesizers. They can not offer learners with learning feedback and high quality of voice. Therefore, our research devises a multimedia English Learning (MEL) system to overcome above two limits, not providing learning feedback and using speech synthesizers.

The Hidden Markov Model (HMM)-based automatic speech recognition (ASR) system has been successfully applied to dictate speech tasks (Nock & Ostendorf, 2003). HMM provides a framework which is broadly used for modeling speech patterns. The hidden Markov model (HMM) is the most commonly employed model for speech recognition. Speech recognition technology based on the HMM using for word spotting and speech recognition, has improved significantly during the past few decades (Liu et al., 2006). Although the HMMs play an important role in most recognition systems for a long time, many alternative models have been proposed in recent years to overcome shortcomings of the HMMs. High recognition rate needs higher pronunciation training cost. A disadvantage of the ASR systems is that they are highly sensitive to variations between training and testing conditions such as changes in speaker voice or acoustic environment. Moreover, a method with hierarchical clustering was proposed (Mathan & Miclet, 1990). However, this method had no depth determination procedure for a word-based speech recognition system (Kosaka, Matsunaga, & Sagayama, 1996). Therefore, the paper proposes an adaptive clustering technique to improve discrepancies mentioned above.

English is regarded as a second language in Taiwan. The learners cannot immediately have learning feedback from teachers and cannot quickly evaluate their learning level. Therefore, English learning performance of learners is not so good. This inspires us to develop a speech learning system to offer the correct feedback to learners and not to sound artificially. Therefore, this paper proposes the MEL system based on HMMs and mastery theory for the learning of English pronunciation. This system adopts a phoneme-based HMMs to perform speech recognition. The system offers feedbacks by integrating a dialogue speech tool for native English pronunciation, phonemic clustering for reducing the computational complexity, and mastery learning theory offers correct feedbacks (Marsha & Marion, 2007). Speech recognition then makes learning evaluation with the four dimensions of pronunciation, intonation, tempo, and volume (Liu et al., 2006). By the aids of this system, the learners can be able to identify their problems for speaking English and make a lot of progress due to the help of different error analyses.

The rest of this paper is organized as follows. Section 2 briefly reviews an ASR system based on HMMs and mastery learning. Section 3 describes the MEL system. Section 4 shows the experiment results, and Section 5 draws conclusions.

Background

The pronunciation difference of EFL learners

Many researches indicate that native language pronunciation significantly affects learning effects for English pronunciation (Jenkins, 2000). EFL learners easily make reading mistakes while they sound English words. The reason is that some sounds of English words are excluded in the set of sounds of native language. Some EFL learners often fuse the intonation and rhythm of the mother tongue into the pronunciation of English language. This causes incorrect pronunciation. Several papers have proposed the results that the mother tongue interferes with pronunciation correctness while speaking foreign language. The pronunciation differences between native language speakers and EFL learners can be summarized as follows (Jenkins, 2000; Wang, 2003).

- Lack: Sounds of some English words do not exist. Therefore, learners are unable to correctly pronounce the words. For example: /æ/ is pounced as /ɛ/ because there is no extremely low-tongue location of pronunciation symbol for native language.
- Substitution: Learners substituted English pronunciation by similar native language. For example: Learners in Taiwan replace /ʃ/ with /tʃ/. This may cause incorrect pronunciation for syllable, intonation, and rhyme.
- Simplification or complexity: Learners often add or omit one consonant due to side effect of speaking mother tongue. For example: "question" is pounced as /kwɛstʃn/.

- Epenthesis: There is no CVC (a Consonant, following Vowel, and then a Consonant) in China's ordinary speech structure. Therefore, some learners may insert one vowel to the last letter of words. Thus, CVC becomes CVCV. For example: “some” /sʌm/ takes place as /sʌmu/.

Mastery learning

The mastery learning is an effective way to make learners reach higher learning level. It aims at that all students can almost reach high levels of competence on instructional material. Bloom (1968) deemed that well organized teaching materials and effectively managing student’s learning process are two effective instruction factors. As conceptualized by Bloom (1976) and others (Block & Burns, 1976; Fuchs, Fuchs, & Tindal, 1986), mastery learning can be accomplished by following procedures. The first step towards the realization of applying mastery learning theory is to divide the concepts and materials into relatively small and sequential learning units. Each unit should be associated with concrete learning objectives. The learning structure is organized by the way form easy units to difficult units. After teaching each unit, a formative assessment is conducted to get the results where the learners have reached the learning level or not, and also to reflect feedback on their learning (Yang & Liu, 2006). The learners, who have not mastered a unit, should enter the process of remedial activities or corrections for fully mastering the unit. The learning process can be shown in Figure 1. Mastery learning is suitable for students due to that they have a weakness in self learning. Therefore, the MEL system applies the mastery learning in the design of learning paths to teach English pronunciation in the rural primarily school in Taiwan.

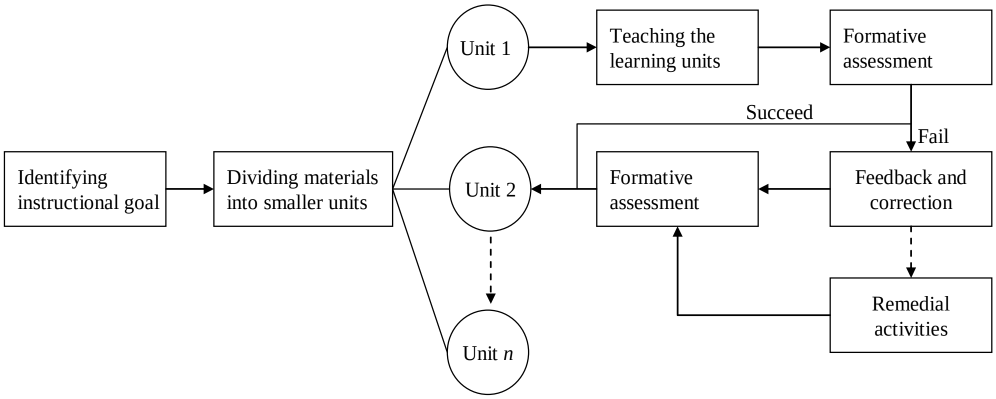


Figure 1. The strategy for mastery learning

A review of an ASR system based on HMMs

The speech signal can be expressed as a form of a sequence of samples (Young et al., 2005). Figure 2 depicts a block diagram of speech recognition using HMMs, which consists of four components, Frame Blocking (FB), Feature Extraction (FE), Parameter Construction (PC), and HMMs. Finally, the recognizer outputs the phoneme of maximum probability to be the recognition result.

Frame blocking

While analyzing speech signal, the frame procedure of blocking is involved first. Frame blocking is to partition a sequence of speech samples into a set of frames. The feature parameters associated with each frame can then be extracted. Figure 3 illustrates an example for frame blocking.

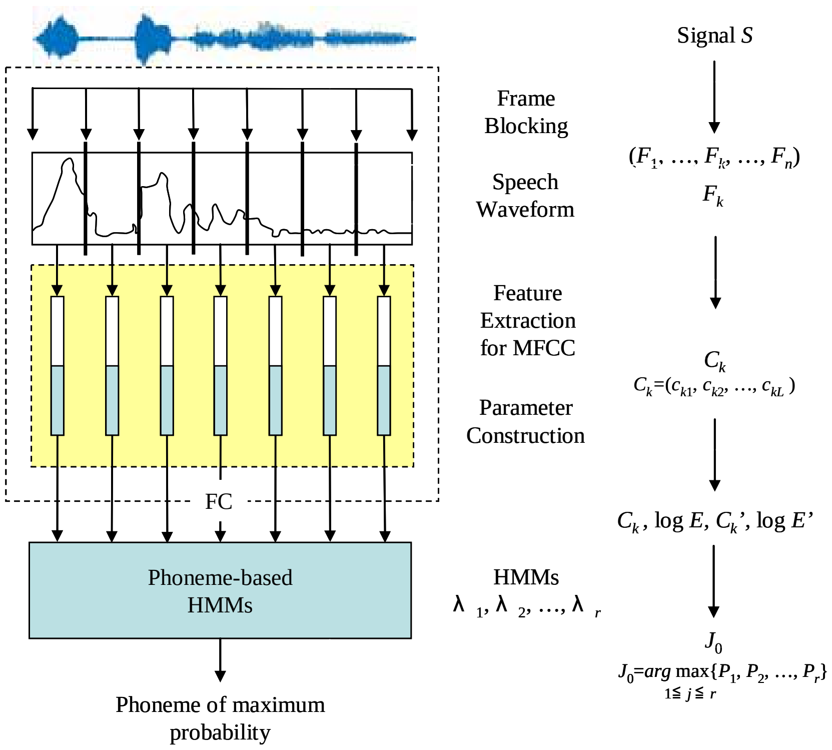

Signal S
Frame Blocking
$(F_1, \ldots, F_k, \ldots, F_n)$
Speech Waveform
F_k
Feature Extraction for MFCC
C_k
$C_k=(c_{k1}, c_{k2}, \ldots, c_{kL})$
Parameter Construction
FC
C_k, log E, C_k', log E'
Phoneme-based HMMs
HMMs
$\lambda_1, \lambda_2, \ldots, \lambda_r$
J_0
$J_0=\arg\max_{1\leq j\leq r}\{P_1, P_2, \ldots, P_r\}$
Phoneme of maximum probability

Figure 2. Block diagram of speech recognition using HMMs

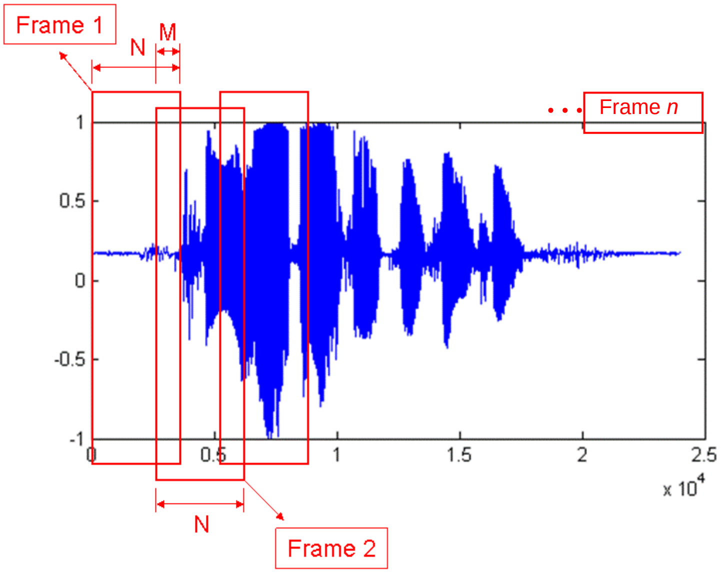


Figure 3. Speech signal with frame blocking

Let a speech signal S be partitioned into a sequence of n overlapped frames, F_1, F_2, …, and F_n, and be represented as

$$S = (F_1, F_2, \ldots, F_n). \tag{1}$$

Assume that the frame duration and the frame overlap for S are N and M, respectively. The first frame F_1 consists of N samples, which can be denoted as

$$F_1 = (x_1, x_2, \ldots, x_N), \tag{2}$$

where x_1 is the first sample and x_N is the Nth sample in F_1. Consequently, the nth frame is specified by

$$F_n = (x_{(n-1)(N-M)+1}, x_{(n-1)(N-M)+2}, \ldots, x_{(n-1)(N-M)+N}). \tag{3}$$

Therefore, the equation (1) can be rewritten as

$$S = (x_1, \ldots, x_i, \ldots, x_{(n-1)(N-M)+N}), \tag{4}$$

where x_i denotes a speech sample, $i = 1, 2, \ldots, (n-1)(N-M)+N$.

In the implementation phase, the frame duration and frame overlap are set to, generally, about 20-30ms (millisecond) and the half of the frame duration, respectively.

Feature extraction

Several kinds of methods can be used to obtain speech feature parameters such as linear prediction coding (LPC), Mel-Frequency Cepstral Coefficients (MFCC), perceptual linear predictive analysis (PLP), etc. The MFCC is the most frequently used in the computation of speech feature parameters (Zheng, Zhang, & Song, 2001). Therefore, the feature parameters, which are calculated by using the MFCC (Figure 4), are especially suitable for the application of speech recognition. After feature extraction, the feature C_k of a frame can be written as

$$C_k = (c_{k_1},...,c_{k_i},...,c_{k_L}, \log E_k, dc_{k_1}, ..., dc_{k_i},..., dc_{k_L}, d(\log E_k)), \tag{5}$$

where E_k represents the energy expression, dc_{k_i} and $d(\log E_k)$ stand for the delta coefficients of c_{k_i} and $\log E_k$, respectively, for i = 1, 2, …, and L. Detail descriptions for MFCC can be found in (Nwe & Li, 2007; Young et al., 2005).

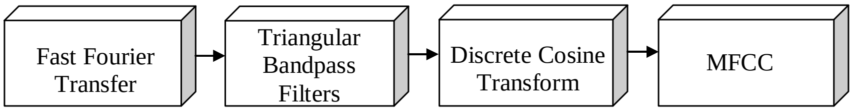


Figure 4. The procedure of feature extraction

HMM

The technique of HMMs has been efficiently applied in speech recognition (Nock & Ostendorf, 2003). Assume an HMM has a set S of M states, $S = \{S_1, S_2, \ldots, S_M\}$. Each state S_i corresponds with a set of transition probability denoted by $A = \{a_{ij} \mid i, j = 1, 2, \ldots, M\}$. Note that a_{ij} denotes the probability of a transformation form state S_i to state S_j. In addition, each state S_i has an observation probability distribution, $B = \{b_j(o_t)\}$, $t = 1, 2, \ldots, T$. Specifically, an observation probabilities is that o_t is observed at state S_j. Accordingly, an HMM can be specified by $\lambda = (A, B, \Pi)$, where $\Pi = \{\pi_i \mid i = 1, 2, \ldots, M\}$, and π_i denotes the initial probability of S_i. Figure 5 exhibits an example of an HMM with M states. An HMM λ_i is used to recognize the phoneme P_i. An HMM λ_i is used to recognize the phoneme P_i. In this paper, 39 HMMs are involved in the design of speech recognition of the MEL system.

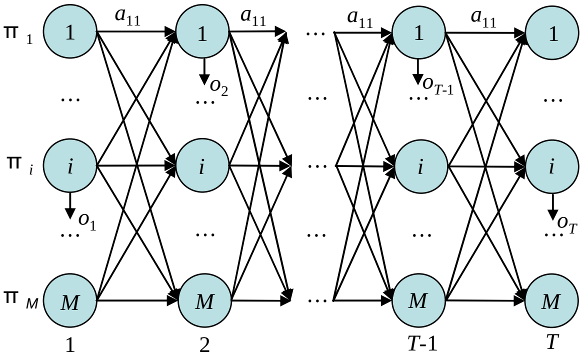


Figure 5. A diagram of HMM with M states

The MEL system

Figure 6 depicts the architecture of the MEL system which consists of three modules, the Speech Analysis Module, the Mastery Learning Module for students, and the Management Module for teachers.

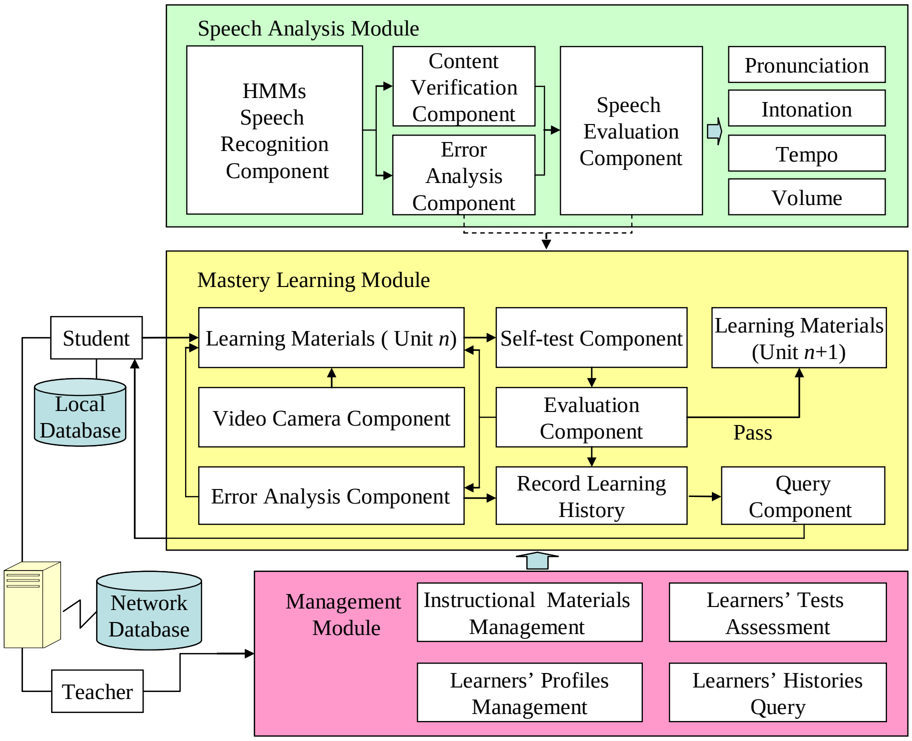


Figure 6. The architecture of the MEL

Speech recognition algorithm of the MEL system

The paper proposes an adaptive phoneme clustering (APC) algorithm. The algorithm is then used in the design of Speech Analysis Module. The goal of the algorithm is to simultaneously reduce the phoneme recognition time complexity and the recognition error rate. The Kenyon & Knott (KK) phonetic symbols are employed in this paper while constructing a phoneme recognition model for each phoneme. Thirty nine phoneme recognition models, which are constructed by HMMs, are shown in Figure 7(a). In order to speed up the recognition process, these phonemes can be classified into clusters to form a hierarchical recognition model with two levels. An example for 5 clusters of the phonemes is exhibited in Figure 7(b).

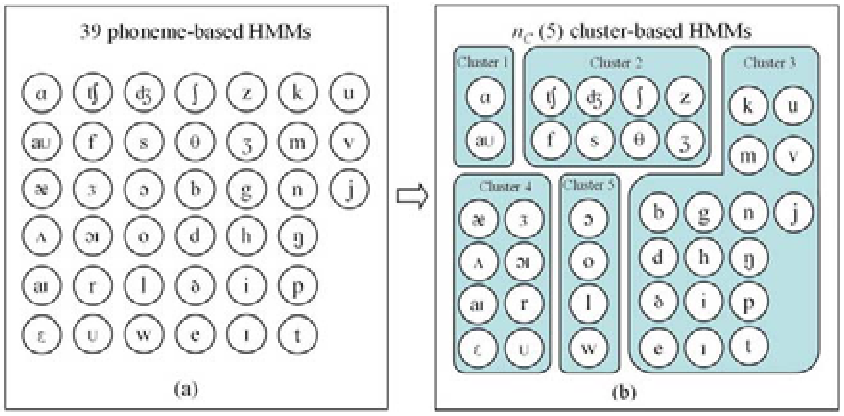


Figure 7. An example of clustering phoneme-based HMMs

During phoneme recognition, the APC algorithm first specifies which cluster the input phoneme belongs to, and then the input phoneme is recognized by involving all HMMs in the specified cluster. This way can reduce the recognition time in contrast to the method recognizing an input phoneme by using all HMMs (39 HMMs). Accordingly, the APC algorithm can reduce the recognition time.

Figure 8 displays a block diagram for the APC algorithm. Assume that there are n_{tr} training patterns, n_{te} testing patterns, and n_c clusters. First, let $\{S_1, S_2, \ldots, S_{n_{tr}}\}$ denote a set of n_{tr} training patterns. After feeding the FE component with S_j, the jth sound signal, each frame F_k^j has a set C_k^j of feature parameters, where

$$C_k^j = (c_{k_1}^j, c_{k_2}^j, \ldots, c_{k_n}^j)\ . \tag{6}$$

Therefore, the training-pattern set can be expressed as a form

$$\zeta = \{\, C_k^j \mid j = 1, 2, \ldots, n_{tr},\ k = 1,2, \ldots, n\}. \tag{7}$$

After feeding the K-means algorithm with the set ζ, a new training-pattern set ζ' can be obtained and represented by,

$$\zeta' = \{\, (C_k^j, \text{cluster}_i) \mid i = 1, 2, \ldots, n_c,\ j = 1, 2, \ldots, n_{tr},\ k = 1, 2, \ldots, n\}, \tag{8}$$

where $C_k^j \in \text{cluster}_i$. In other words, ζ' is employed to construct a classification model which is realized by HMMs.

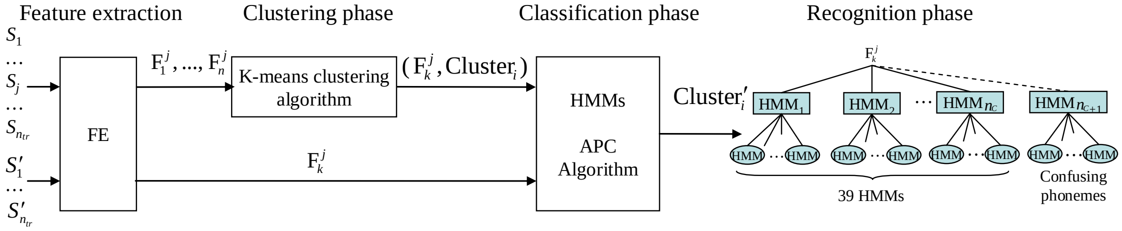


Figure 8. A block diagram for the APC algorithm

Since the Euclidean distance between two feature vectors C_k^j does not effectively measure the similarity between a phoneme and others, the HMMs are applied to model the phoneme recognition instead of applying the *K*-means algorithm to perform phoneme recognition. Table 1 presents an example of using the APC algorithm for the case of two clusters. The l-phoneme in the bottom row in Table 1 should be in Cluster 1 but it is labeled with Cluster 2 by using *K*-means algorithm. Note that in this paper the training data is the TIMIT speech database (Garofolo & Lamel, 1993).

Confusing phoneme

Some phonemes cannot be recognized correctly because these phonemes apparently exist in more than one cluster. In order to avoid incorrect cluster recognition leading to lower accuracy rate of phoneme recognition, the APC algorithm specifies a set of confusing phonemes, and then reconstructs the 2-level HMMs model with (n_c+1) clusters. More specifically, ζ' is extended to ζ'' by replacing (C_k^j, cluster$_i$) with (C_k^j, cluster$_{n_c+1}$) if C_k^j is a feature of a confusing phoneme. Then, a 2-level HMMs with (n_c+1) clusters can be obtained by feeding the n_c-HMMs algorithm with ζ''. A phoneme is considered as a confusing phoneme if its error rate, $\pi_{(2)} / \sum_{i=1}^{n_c} \pi(P, i)$, exceeds a threshold, where $\pi_i = \pi(P, i)$ denotes the number of phoneme P in the Cluster$_i$, and $\pi_{(1)} \geqq \pi_{(2)} \geqq \ldots \geqq \pi_{(n_c)}$. Note that the set of clusters is produced by using *K*-means algorithm. An example given in Table 2, there are two confusing phonemes, w and h if the threshold is set to 0.1. Table 3 shows the recognition accuracy of the MEL system in comparison with other existing methods such as Dynamic HMM (Salmela, 2001), Factorial-HMM (Virtanen, 2006), and PT-FHMM (Nock & Ostendorf, 2003).

Table 1. An example of phoneme clustering adaptive method with two clusters

Test Data	Test Data	*K*-means	HMMs	Experimental Results	
(Code)	(KK)	algorithm		Cluster 1	Cluster 2
AA	ɑ	1	1	107	1
EH	ɛ	1	1	145	1
ER	ɝ	1	1	189	3
IH	ɪ	1	1	352	12
HH	h	2	2	21	55
W	w	2	2	24	70
L	l	**2**	**1**	**134**	**39**

Table 2. An example of confusing phoneme with two clusters

Test Data	Test Data	*K*-means Cluster	Experimental Results		
(Code)	(KK)		Cluster 1	Cluster 2	Error Rate
AA	ɑ	1	107	1	0.009
AE	æ	1	158	1	0.006
AH	ʌ	1	171	4	0.023
AY	aɪ	1	100	0	0
ER	ɝ	1	189	3	0.016
W	w	**2**	**24**	**70**	**0.255**
HH	h	**2**	**21**	**55**	**0.276**

Table 3. The recognition accuracy of the MEL system in comparison with other existing methods

Type	Dynamic HMM	Factorial-HMM	PT-FHMM	MEL
Accuracy	96.8%	97.0%	97.4%	97.5%

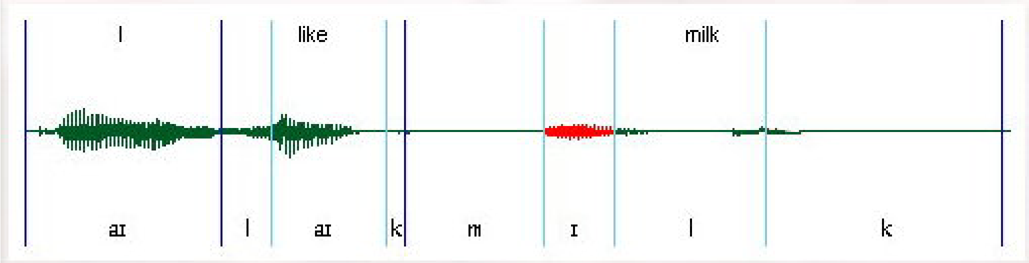


Figure 9.(a) The pronunciation error

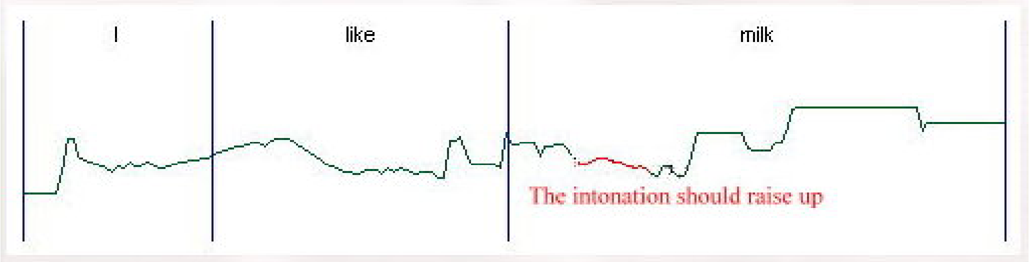


Figure 9.(b) The intonation error

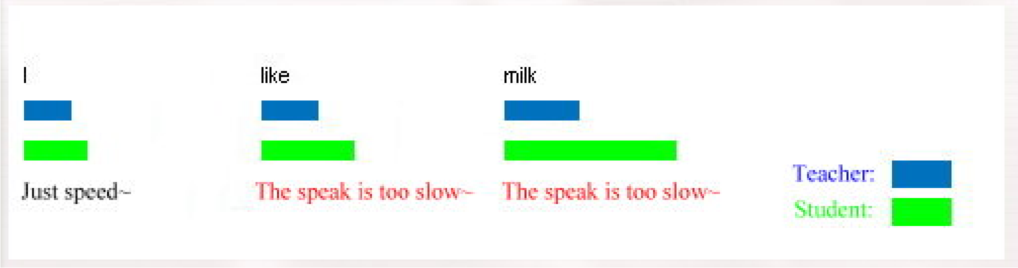


Figure 9.(c) The tempo error

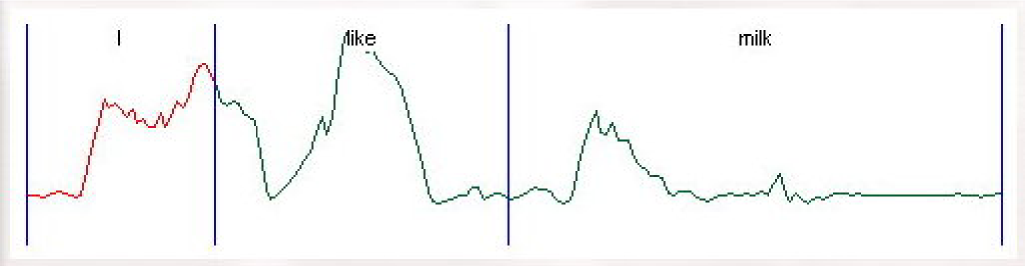


Figure 9.(d) The volume error

The models of the MEL system

Speech analysis module

In the MEL system displayed in Figure 6, the Speech Analysis Module is used to analyze the speech signal and to detect the correction of student’s pronunciation. The HMMs Speech Recognition Component is employed to recognize the input speech signal and also computes the probability and ranking of phonemes. The Content Verification Component is used to prevent random utterances and the Error Analysis Component is used to compare student’s pronunciation with the standard version. The Speech Evaluation Component then evaluates the correctness of the input speech signal in terms of the four factors: pronunciation, intonation, tempo, and volume.

For the pronunciation, when a student reads a sentence, the system can clearly show his/her incorrect pronunciation for the word, phrase, or sentences. It marks the interval of the voice signal to analyze the pronunciation error if the pronounce score is less than 60. Figure 9(a) shows an interval of the pronunciation error colored in Red. For the intonation, the MEL system provides the intonation curves of the teacher with students. It can display incorrect intonation to help students to correct their intonations. Figure 9(b) exhibits an interval of the intonation error colored in Red. For the analysis of tempo, the MEL system can compare the voice speed at which every student reads every word, phrase, and sentences with teachers. Figure 9(c) displays an interval of the tempo error. For the analysis of volume, the MEL system compares the volume of students with teacher’s voice for their volume stress. Figure 9(d) shows an interval of the signal at aloud volume.

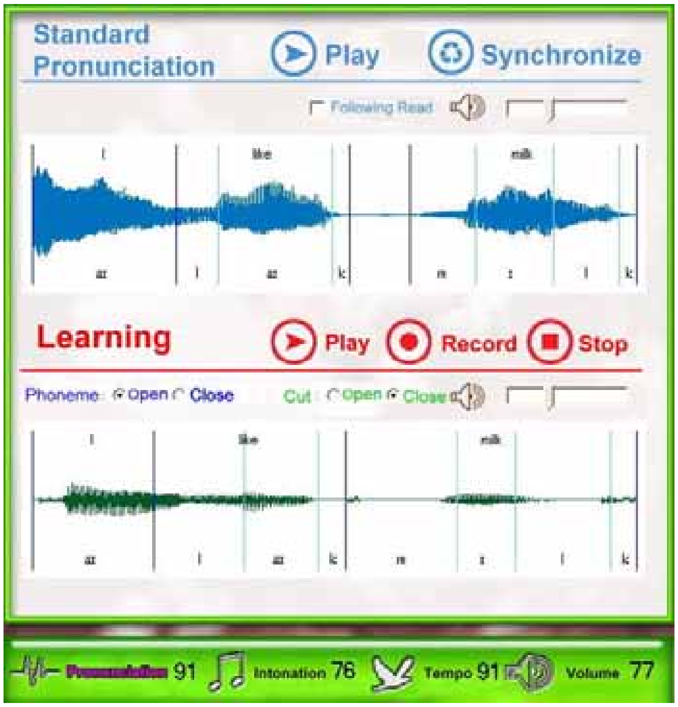
Mastery learning module for students

The students’ learning module includes the Learning Materials Component, Self-test Component, Error Analysis Component, Video Camera Component, Record Learning History, and Query Component. The module supports small-scale teaching, sufficient opportunities to practice, plenty time to learn, and feedback for the learning. Students can practice in the units of Learning Materials Component. Each unit is divided into words, idioms and phrases, and sentences. Students can decide the learning sequence for unit’s content.

The Self-test Component is used to test the student when s/he complete study from current unit. Evaluation Component is employed to support the standard pronunciation to student and decides the standard score for which student can pass the unit. Students can pass the unit if they get a score of 85. Once they pass the test, they can continue to the next unit.

According to the results of the speech analysis module, the Error Analysis Component is used to offer students with the error analysis for the pronunciation, intonation, tempo, and volume. After practicing or testing, students can get a feedback from the view to watch their sound waves and teacher’s sound waves on the screen. There are three switch buttons for intonation, tempo, and volume. Students can get the performing waves and look out for errors in each of the four factors. Students can compare their results with teachers (standard sentences) to obtain the errors in terms of these four factors. Figures 9-10 show an example for student’s sound wave of the sentence “I like milk”. Students can see the individual score and errors for each question and, therefore, they can know the exact problems such as the pronunciation error, the intonation error, the tempo error, and the volume error while they practice their speech. Students can obtain the adequate feedback on their spoken English and naturally improve it. Video Camera

Component with charge-coupled device (CCD) function can be used to record students' lips. By replaying video, the student can review the mouth shape s/he makes while speaking.



Pronunciation 91 Intonation 76 Tempo 91 Volume 77

Figure 10. Student's sound wave of the sentence "I like milk"

The system can show the corresponding relation of shape-phoneme to enhance the ability of pronunciation. Figure 10 shows the corresponding of letters and phonetic symbols for student's phonemic awareness. The instruction strategy mainly relies on spelling and combining the syllables in letters, emphasizes goal analysis, word of phoneme. Students may strengthen their phonemic awareness by dividing each phoneme into a segment for each speech sound in a word. For example, while students are pronouncing the word "milk", they pronounce "milk" as the /mɪlk/ sound. In practice, students first pronounce the letter "m" as /m/ sound, the letter "i" as /ɪ/ sound, the letter "l" as /l/ sound, and finally the letter "k" as /k/ sound. Subsequently, the students can regard letter "m" as prefixes to practice the associating words whose the first letter is "m", such as "mill". This way can increase the practice for the pronunciation to letter "m". The students replace the second letter "i" with letter "e", such as "melk". This way can increase the practice for the pronunciation to the second letter "i" and "e".

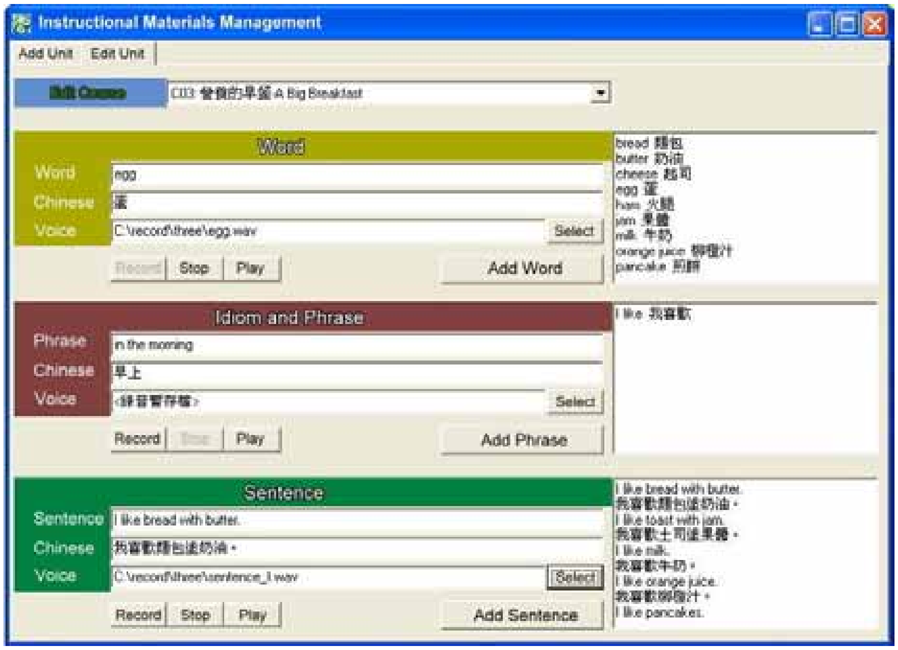
The component, Record Learning History, is designed to record full learning experience of each student progress and test for each unit. Students can use Record Learning History to obtain their learning level. Query Component is designed to query the learning history of each unit which a student has completed the test. Students can use Query Component to get their individual score for each question and to understand the problems they have with their speech. Therefore, students can obtain adequate feedback on their spoken English and naturally improve their pronunciation skills through the MEL system which offers the students with learning feedback to modify their learning paths.

Management module for teachers

The teacher's management module can be used to manage instructional materials, manage students' profiles, assess students' tests, and query students' learning histories. The functions of management module are as follows.

The Instructional Materials Management contains the Add Unit and the Edit Unit. Teachers utilize the Add Unit while creating a new unit. Teachers can add or edit words, idioms and phrases, and sentences as new materials by using the Edit Unit. The Edit Unit also offers users a record function to record teachers' or experts' own voice which can be regarded as teaching and learning resources. Figure 11 shows the Instructional Materials Management for editing the learning unit.

Teachers employ the Learners' Profiles Management to add and edit students' profiles. Teacher can add and manage students' accounts. The Learners' Tests Assessment is used to offer teachers to edit the questions or to set the numbers and types of questions which the test of each unit will appear. The Learners' Histories Query is used to provide teachers with a function to query students' learning histories, such as the test scores of each student and detailed score distributions. Therefore, teachers can obtain students' learning status and then provide them with appropriate assistances or guidances.



Edit Course: C03 營養的早餐 A Big Breakfast

Word: egg
Chinese: 蛋
Voice: C:\record\three\egg.wav

bread 麵包
butter 奶油
cheese 起司
egg 蛋
ham 火腿
jam 果醬
milk 牛奶
orange juice 柳橙汁
pancake 煎餅

Phrase: in the morning
Chinese: 早上
Voice: <錄音暫存檔>

I like 我喜歡

Sentence: I like bread with butter.
Chinese: 我喜歡麵包塗奶油。
Voice: C:\record\three\sentence_1.wav

I like bread with butter.
我喜歡麵包塗奶油。
I like toast with jam.
我喜歡土司塗果醬。
I like milk.
我喜歡牛奶。
I like orange juice.
我喜歡柳橙汁。
I like pancakes.

Figure 11. The Instructional Materials Management for editing the learning unit

Experiment and results

Participants

In Taiwan, English is the second language. Students at the third grade begin to learn English listening and speaking in elementary school. A total of 120 third-grade students (67 females and 53 males) from an elementary school in Yunlin County participated in the study. They were recruited and randomly assigned as the experimental group and the control group, respectively. Students were ranged in age from 9 to 10 years (*Mean* = 9.6). Based on the mastery theory, the main purpose of this study was to make most of students mastery the learning contents. Therefore, this study used the phonemic awareness scores to classify students into three categories: (a) high-score group including the top 27% of the samples (experimental group *n* = 16; conventional group, *n* = 16), (b) middle-score group including the middle 46% (experimental group, *n* = 28; conventional group, *n* = 28), and (c) low-score group including the bottom 27% (experimental group, *n* = 16; conventional group, *n* = 16).

Assessment materials

Phonemic awareness test (PAT)

Phonological awareness, a total of 50 items, was measured by using the three subtests to assess phonemic recognition, deletion, and segmentation. First, the process of phonemic recognition subtest was that student pointed out which word in a list of words while s/he heard the sound of the target word. E.g., teacher pronounces the "joak" in the list “joap, joak”, and then student should point out “joak” as the correct response. The participants were asked to recognize 16 lists with two printed words or pseudowords in each list. Subsequently, the phonemic deletion subtest was that the student deletes the initial, medial, or final phoneme of a new word or pseudoword. Student was instructed to pronounce an original word and then asked to pronounce the deleted word after removing a specific phoneme of the word. E.g., the original word was “boat”. Teacher deleted the initial “b” and then the student was asked to pronounce the deleted word “oat” without teacher’s instruction. The phonemic deletion subtest score was measured with 18 items from the task with 6 initial, 6 middle, and 6 final deletions (Leong, Tan, Cheng, & Hau, 2005). Finally, the phonemic segmentation subtest was modified by the vowel phoneme-grapheme correspondence (Landerl & Wimmer, 2000). Student listened the word-item pair and replaced the first grapheme, e.g., the words “boat-c” and “chair-t” were replaced with “coat” and “tair”, respectively. There were 16 words used in the subtest. The internal reliability of Cronbach's coefficient α for this test was .86.

Learning achievement test (LAT)

Learning achievement test consisted of two tasks of spelling (30 points) and reading (30 points). In the first task, the spelling subtest consisted of 30 words where each word was selected from a sentence. In addition, those 30 words were sequenced in order of difficulty. E.g., teacher read the sentence “I have a lunch *break*” then asked student to spell the assigned word “*break*” with “b” “r” “e” “a” “k” (Kaufman & Kaufman, 1985). In this sample, the internal reliability of Cronbach's coefficient α was .86. In the second task, every student read the reading Subtest I which includes 20 words in an increasing difficult order. After a student sounds the 20 words, the teacher will score the subtest for the student (Sullivan & Hawkins, 1995). The internal reliability of Cronbach's coefficient α was .89. The reading Subtest II which includes 5 questions was used to evaluate how well students comprehend what they had read in learning units. For example, one question “Do you like salad?” four choices were given in the following (a) She is old, (b) She like it, (c) Not very much, or (d) Thank you!

Procedure

This experiment lasted 12 weeks to teach the third graders. Students were taught in their normally scheduled English language classes with two 40-min class periods a week. During the preparatory activities, researchers and teachers agreed to select six units ‘Early in the morning’, ‘I like noodles’, …, and ‘The end of the day’ from the text book. Each unit in the teaching activities includes warm up, review, vocabulary, pattern, chant, dialogue, and assignment. For the activity of vocabulary and pattern, teachers use the MEL system to teach students to learn words, idioms, phrases, and sentences. Meanwhile, teachers also teach students how to use the MEL system after class. The system can show the corresponding relation of shape-phoneme to enhance the ability of pronunciation. Figure 9 shows the corresponding of letters and phonetic symbols for student’s phonemic awareness. The instruction strategy mainly relies on spelling and combining the syllables in letters, emphasizes goal analysis, word of phoneme. Students may strengthen their phonemic awareness by dividing each phoneme into a segment for each speech sound in a word. The practice is as follows.

1. Teachers record their or the native speaker pronunciations.
2. Students utilize the “play key” in the MEL system to listen to pronunciations recorded in the database of the MEL system.
3. Students loudly pronounce vocabularies, idioms, phrases, and sentences, and then the system records students’ pronunciations.
4. The system compares students’ pronunciation with that of the teachers for the four factors, the pronunciation, the sound intonation, the tempo, and the volume.
5. The system provides students with the learning results as shown in Figures 9-10.
6. According to the results, students obtain whether their pronunciations are well or poor.
7. Students repeat the practices to correct their errors if students deem their pronunciations have to be improved.
8. Students also practice similar words or similar sounds to make students easily obtain comparison results for the sound characteristics and differences among these similar words or similar sounds.
9. Teachers can understand the student’s learning performance by analyzing the learning profile of each student.

Before teaching, all students took Phonemic Awareness Test to compare the posttest with pretest after teaching. In the instructional activities, the teacher explained to students how to use the MEL system to proceed with their learning. Also, the teacher used the MEL system only to assist and instruct the experimental group, so as to supply students with vocabulary listen, vocabulary presentation, pattern presentation, rhyme presentation, and test assignment. After teaching, students in the experimental group also were able to study with the MEL system and students in the control group only studied with the traditional method after the class. After 12 weeks, all students took Phonemic Awareness Test and English Achievement Test.

Data analysis

Research data were collected through the tests of the Phonemic Awareness Test and English Achievement Test. The experimental group adopted the MEL system in English learning for phonemic awareness, while the control group received the conventional English teaching and learning. Independent-samples *t*-test was involved to compute the values of the means on the phonemic awareness scores to examine for differences in pretest and posttest between the experiment group and control group. Here, the significant level was set at $p = 0.05$. The two-way ANOVA test with 2 (instructional method: MEL and convention) × 3 (phonemic awareness level: high, middle, and low) factorial design was applied to investigate the differences between the MEL and convention for students' learning achievement.

Results

The MEL system involves phonemic awareness instead of a word. It applies mastery theory to design learning process to effectively reach learner's achievement. It provides learners with feedback consist of four features, pronunciation, intonation, speed, and volume. Table 4 shows the characteristics of the MEL system in comparison with other similar methods such as iKnowthat (iKnowthat, 2007), Phonetics Flash Animation Project (Library of English sounds, 2005), English Pronunciation (College English Web, 2006), and FluSpeak (MT Comm, 2002).

Table 5 shows the results that the differences in posttest of the experimental group with the control one, we had statistically significant differences between the two groups, $t(118) = 2.489$, $p < .05$. It was discovered that the students who studied with the MEL system obtained, on average, a better result. The mean of scores was 72.80 (*SD* = 17.35) for the experimental group, higher than the 64.55 (*SD* = 18.93) for the control one. The results showed that the experimental group was more effective than the control one.

A 2 (instructional method) × 3 (phoneme) ANOVA revealed that MEL students' LAT scores were higher than conventional students' scores, $F(1, 114) = 11.83$, $p < .01$, $\eta^2 = .09$, and students' achievement test scores varies directly with phoneme, $F(2, 114) = 21.57$, $p < .01$, $\eta^2 = .27$ (Table 6). Scheffe's post-hoc test ($p < .05$) indicates that all three phoneme groups differed from one another. The instructional method by phoneme interaction is not significant, $F(2, 114) = 0.686$, $p > .05$, $\eta^2 = .012$. Independent *t* test reveals that the MEL students at low and middle phoneme level scored higher on the LAT than conventional students: low-phoneme group, $t(30) = 2.56$, $p < .05$; middle-phoneme group, $t(54) = 2.80$, $p < .05$; but not higher at high- phoneme group, $t(30) = 0.97$, $p > .05$. The MEL students scored higher, 3.62 (6.03%), in the low-phoneme group and, 3.53 (5.88%), in the middle-phoneme group on the LAT compared to conventional students. In contrast, the advantage of the MEL system is only 1.44 (2.4%) for the high-phoneme group.

Table 4. The characteristics of the MEL system in comparison with other similar methods

Methods Items	iKnowthat	Phonetics Flash Animation Project	English Pronunciation	FluSpeak	MEL system
Pronunciation	Yes	Yes	Yes	Yes	Yes
Intonation	No	Yes	Yes	Yes	Yes
Speed	No	No	No	No	Yes
Volume	No	Yes	No	Yes	Yes
Scoring	Yes	No	No	Yes	Yes
Using HMMs	No	No	No	Yes	Yes
Materials management	No	No	No	No	Yes
Manage learning process	No	No	No	No	Yes
Split a word into phonemes	No	No	No	No	Yes

Table 5. The number (*n*), means (*M*), standard deviations (*SD*), and t value on the pre- and post-test with the ability of phonemic awareness

	MEL			Convention			*t* value
	n	*M*	*SD*	*n*	*M*	*SD*	
Pre-test	60	51.97	24.34	60	49.62	27.19	0.499
Post-test	60	72.80	13.75	60	64.55	18.93	2.489*

* $p < .05$; ** $p < .01$; *ns* = not significant

Table 6. Learning achievement test scores as a function of instructional method and phoneme

	MEL			Convention			Total		
Phoneme	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
High	16	45.19	4.23	16	43.75	4.19	32	44.47	4.20
Middle	28	42.21	4.35	28	38.68	5.06	56	40.45	5.01
Low	16	39.06	3.71	16	35.44	4.29	32	37.25	4.36
Total	60	42.17	4.67	60	39.17	5.52	120	40.67	5.31

* $p < .05$; ** $p < .01$; *ns* = not significant

Discussion and conclusion

Phonemic awareness is an important meta-linguistic skill which can let students more effectively acquire reading and spelling abilities. While children learn English, an important step is to train them with high phonemic awareness. The paper has presented the MEL system which analyzes audio samples from English-learning students, compares the student's samples with those samples of native speakers or teachers, and evaluates whether the pronunciation is correct or not. The phonemes of the samples are analyzed according to four factors, pronunciation, intonation, rhythm, and volume. The APC method is employed in the design of the MEL system for reducing the computation time of the hierarchical HMMs. The system utilizes HMMs to obtain phonemic features which are subsequently used in the process of English pronunciation errors. According to the pronunciation errors, the system provides students with advices in these four criterions for students to correct and improve their pronunciations and to improve learning effects.

Our findings reflect that the MEL system can promote the phonemic ability of the students with the middle and the low phonemic ability. These findings also support that mastery learning makes low-ability students to devise an effective control over learning situations and more opportunities in English learning courses. This concludes that the MEL system improve student's mastery level for learning and help them to obtain more achievement for English pronunciation learning. From the pedagogical point of view, possible impacts of this study are summarized as follows. First, teacher can employ the MEL system to quickly obtain student's pronunciation results (errors) in terms of four factors: pronunciation, intonation, rhythm, and volume. Accordingly, this way is easier to realize the elaborating instruction than traditional pronunciation in classrooms (without the MEL system). Second, when students repeatedly practice pronunciations, the MEL system can interactively provide concrete feedbacks. It is helpful for self-regulated learning (Butler & Winne, 1995; Mondi, Woods, & Rafi, 2008). Third, the MEL system can be readily incorporated into e-Learning environments to perform asynchronous learning so that students can practice pronunciation from anywhere at anytime.

Acknowledgements

We thank the anonymous reviewers for their constructive comments and the National Council of Taiwan under Contract Nos. NSC 96-2520-S-194-002-MY3 and NSC 96-2221-E-150-062.

References

Block, J. H., & Burns, R. B. (1976). Mastery learning. In L. S. Shulman, (Ed.). *Review of Research in Education, 4*, 3-49. Itasca, IL: Peacock.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1-5.

Bloom, B. S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Carreker, S. (2005). *Teaching spelling.* In J. Birsh (Ed.), Multisensory teaching of basic language skills (2nd Edition), 257-295. Baltimore, Md.: Paul Brookes.

College English Web. (2006). *English Pronunciation.* English Language Laboratory, Fu Jen Catholic University. Retrieved March 27, 2007, from http://www.lage.fju.edu.tw/sunny/index2.htm.

Fuchs, L. S., Fuchs, D., & Tindal, G. (1986). Effects of mastery learning procedures on student achievement. *Journal of Educational Research, 79*(5), 286-291.

Garofolo, J. S., & Lamel, L. F. (1993). *DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CD-ROM.* USA, Department of Commerce.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL, 15*(1), 3-20.

iKnowthat. (2007). *iKnowthat.* Pittsburgh, PA. Retrieved May 16, 2007, from http://www.iknowthat.com/com/L3?Area=L2_LanguageArts.

Jenkins, J. (2000). *The phonology of English as an international language.* Oxford: Oxford University Press.

Kosaka, T., Matsunaga, S., & Sagayama, S. (1996). Speaker-independent speech recognition based on tree-structured speaker clustering. *Computer Speech and Language, 10*(1), 55-74.

Kaufman, A. S., & Kaufman, N. L. (1985). *Kaufman test of educational achievement.* Circle Pines, MN: American Guidance Service.

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics, 21*(2), 243-262.

Leong, C. K., Tan, L. H., Cheng, P. W., & Hau, K. T. (2005). Learning to read and spell English words by Chinese students. *Scientific Studies of Reading, 9*(1), 63-84.

Library of English sounds. (2005). *Phonetics Flash Animation Project.* Iowa, USA. Retrieved June 18, 2007, from http://www.uiowa.edu/~acadtech/phonetics/english/frameset.html

Liu, Y., Chawla, N. V., Harper, M. P., Shriberg, E., & Stolcke, A. (2006). A study in machine learning from imbalanced data for sentence boundary detection in speech. *Computer Speech and Language, 20,* 468-494.

Mann, V. A., & Foy, J. G. (2007). Speech development patterns and phonological awareness in preschool children. *Annals of Dyslexia, 57*(1), 51-74.

Marsha, I., & Marion, A. (2007). Mastery learning benefits low-aptitude students. *Teaching of Psychology, 34*(1), 28-31.

Mathan, L., & Miclet, L. (1990). Speaker hierarchical clustering for improving speaker-independent HMM word recognition. *In Proceedings of ICASSP'90 Acoustics, Speech, and Signal Processing* (pp. 149-152), Albuquerque, NM, USA.

Mehta, P., Foorman, B. R., Branum, M. L., & Taylor, P. W. (2005). Literacy as a unidimensional construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading, 9*(2), 85-116.

Metsala, J. L. (1999). Young children's phonological awareness and non-word repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*(1), 3-19.

Mondi, M., Woods, P., & Rafi, A. (2008). A ‘uses and gratification expectancy model’ to predict students’ ‘perceived e-learning experience’. *Educational Technology & Society, 11*(2), 241-261.

Morrow, L. M., & Tracey, D. H. (1997). Strategies used for phonics instruction in early childhood classrooms. *The Reading Teacher, 50*(8), 644-651.

MT Comm. (2002). *FluSpeak* (*ASR Software*), Seoul, Korea. Retrieved March 2, 2007, from http://www.fluspeak.com/

Neumeyer, L., Franco, H., Digalakis, V., & Weintraub, M. (2000). Automatic scoring of pronunciation quality. *Speech Communication, 30*(2-3), 83-93.

Nock, H. J., & Ostendorf, M. (2003). Parameter reduction schemes for loosely coupled HMMs. *Computer Speech and Language, 17*(2-3), 233-262.

Nwe, T. L., & Li, H. (2007). Exploring vibrato-motivated acoustic features for singer identification. *IEEE Transactions on Audio, Speech, and Language Processing, 15*(2). 519-530.

Precoda, K., Halverson, C. A., & Franco, H. (2000). Effects of speech recognition-based pronunciation feedback on second-language pronunciation ability. *In Proceedings of InSTILL 2000* (pp. 102-105), Dundee, UK.

Salmela, P. (2001). Applying dynamic context into MLP/HMM speech recognition system. *Computer Speech and Language, 15,* 233-255.

Shinji, S., Keiichi, T., Takashi, M., Takao, K., & Tadashi, K. (2002). Spoken language processing and applications. HMM-based audio-visual speech synthesis. Pixel-based approach. *Transactions of Information Processing Society of Japan, 43*(7), 2169-2176.

Sullivan, T. E., & Hawkins, K. A. (1995). Support for abbreviation of the wide range achievement test-revised spelling subtest in neuropsychological assessments. *Journal of Clinical Psychology, 51*(4), 552-554.

Treiman, R., & Baron J. (1983). Phonemic-analysis training helps children benefit from spelling-sound rules. *Memory and Cognition, 11*(4), 382-389.

Virtanen, T. (2006). Speech recognition using factorial hidden Markov models for separation in the feature space. *In INTERSPEECH 2006 – ICSLP, Ninth International Conference on Spoken Language Processing,* (pp. 173-176), Pittsburgh, PA, USA.

Wachowicz, K. A., & Scott, B. (1999). Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal, 16*(3), 253-276.

Wang, C. L. (2003). From English orthography-phonology correspondence to phonics instruction. *Tunghai Journal of Humanities, 44,* 280-308.

Yang, H. L., & Liu, C. L. (2006). Process-oriented e-learning architecture in supporting mastery learning. *International Journal of Innovation and Learning, 3*(6), 635-657.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (2005). *The HTK Book (for HTK Version 3.3).* Microsoft Corporation.

Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. *Journal of Computer Science and Technology archive Volume, 16*(6), 582-589.

Zinovjeva, N. (2005). Use of speech technology in learning to speak a foreign language. *Term paper, Speech Technology.* Retrieved March 22, 2007, from://www.speech.kth.se/~rolf/NGSLT/gslt_papers_2005/ Natalia2005.pdf.