Linguistic Support of a CAPT System for Teaching English Pronunciation to Mexican Spanish Speakers

Olga Kolesnikova

Superior School of Computer Sciences (ESCOM), Instituto Politécnico Nacional, Mexico City, 07738, Mexico

kolesolga@gmail.com

Abstract. This paper presents an error prevention approach in the development of a Computer Assisted Pronunciation Training System (CAPT System) for teaching American English pronunciation to adult Mexican Spanish speakers developed for the case of vowels. Prior knowledge of the learner's typical articulatory and auditory perception errors enhances the effectiveness of training. It enables to organize the teaching material in a way that foresees possible errors as well as to select appropriate exercises instead of using general pronunciation drills. Our first contribution is an extended comparative analysis of American English and Mexican Spanish vowels at the level of both phonemes and allophones. To the best of our knowledge, such analysis was not done in previous work. Another contribution of this work is a two-fold application of our analysis results: on the one hand, we use the results in designing error patterns to be employed in the error detection module of the CAPT system, and on the other hand, we employ the results as a basis for creating the linguistic content of the error prevention oriented tutor module of the same system. We illustrate this approach with examples.

Keywords. Teaching of pronunciation, English, Mexican Spanish speakers.

1 Introduction

I can speak, read and write but do not understand what they say! So many times EFL (English as a Foreign Language) teachers heard this perhaps most typical complaint of practically every EFL learner. The main reason of this problem is the absence of some English sounds in the learner's mother tongue or significant differences between the sound systems of English and L1 (First Language). This causes errors in sound recognition, and as a result, a considerable lack of understanding.

One of the means to resolve the problem of auditory phonemic recognition failure is an adequate teaching of English pronunciation since auditory comprehension is tightly connected with the human articulatory skills and phonemic knowledge as shown by research of phonological development in infants [28] and the development of listening/reading comprehension of children in the early grades [31]. These results can be applied to adult English learners since they follow similar language acquisition stages as children learning L1 [10]. Also, it is suggested that phonemic awareness,

and consequently listening comprehension, depends on articulation accuracy alongside with other factors like vocabulary size, topical knowledge, psychological and social aspects [24]. In this work, we focus on teaching pronunciation with the objective to improve listening comprehension suggesting that an EFL student has to get familiarized with English sounds, learn how to articulate them and after that get an appropriate training on phonemic, lexical, and phrase auditory recognition. Specifically, we argue that L1-oriented explanation of English sounds and language therapy based exercises can improve the overall quality of language learning.

This paper describes principles and gives examples of presenting and explaining the articulation of English sounds as well as exercises for the training stage. These principles and exercises are based on comparative articulatory phonetics for the explanation stage and speech and language therapy (SLT) for the training stage. In this work, we chose Mexican Spanish and American English language pair. The underlying idea in our work is that error prevention based on identified language-dependent error patterns is a more effective approach than error correction. It is not only efficient and emotionally comfortable for the learner, but also more feasible to implement in the pronunciation error detection module of CAPT systems, since at present automatic error detection irrespective to L1 has not yet reached a high quality level due to computational complexity of automatic speech recognition (ASR) task, see more detail in an overview of CAPT applications, Section 2.

Comparing phonetic systems of American English and Mexican Spanish helps us to predict specific articulation and recognition difficulties which may be experienced by Mexican Spanish ELT learners due to the assimilation effect. Based on such predictions, we develop error patterns and target phoneme presentations which anticipate the learner's problems in English pronunciation. This way the learner will comprehend a very important fact that pronunciation errors are not only "something that is not correct" but they are natural and normal steps in acquiring English. By means of errors, their adequate comprehension and more specific corrective training, the learner develops new articulatory and auditory habits. Such approach creates a stress-free context and helps learners to get rid of the fear of committing an error which does not allow them to make progress in language learning. Such fear is more typical in adult learners than in children. Besides, viewing errors as a "speech disorder", i.e., considering English sounds as correct and the corresponding Mexican Spanish assimilations as incorrect, we can apply the SLT techniques to deal with such "disorders".

Another consideration on pronunciation training is worth mentioning here. We deliberately use the term "English sounds" instead of "English phonemes" because we suggest that part of the reason of pronunciation and auditory comprehension problems is insufficient training, and sometimes an absence of training, in English allophones. If only phonemes are taught, the learner then is not able to recognize them when exposed to allophones which differ significantly from the "classical" phoneme presentations. Therefore, comparing American English and Mexican Spanish sound systems, we deal with the basic allophones.

The rest of the paper is organized as follows. Section 2 gives a brief overview of CAPT systems with a special attention to pronunciation assessment and error detection as well as a summary of existing ESL materials for Spanish speaking learners. Section 3 presents the basic architecture of a CAPT system. Section 4

contains a detailed comparative analysis of American English and Mexican Spanish vowel system on the level of both phonemes and allophones. Sections 5 and 6 list and explain error patterns for the error detection module of the system. Examples of teaching American English vowel sounds on the basis of comparative phonetic analysis are given in Section 7. Finally, we present conclusions and outline future work in Section 8.

2 Related work

2.1 CAPT Systems

Today it is practically beyond doubt that Computer Assisted Language Learning (CALL) is able to provide many benefits to teachers and learners including stress-free and interaction-rich context where teachers enjoy more opportunity to attend individual needs of students, since not all situations can be previsioned and programed in a computer application, while the students can practice at their own pace and get immediate personalized feedback [13]. Besides, techniques of electronic [11], mobile [14] and ubiquitous [2] learning further increase the effectiveness of acquisition.

The majority of CALL applications are oriented to acquisition of all language aspects: phonetic system, lexicon and word usage, grammar, pragmatics. However, a number of systems have been designed for Computer Assisted Pronunciation Training (CAPT), including the following commercial products: *NativeAccent*™ by Carnegie Mellon University's Language Technologies Institute, www.carnegiespeech.com; *Tell Me More*® Premium by Auralog, www.tellmemore.com; *EyeSpeak* by Visual Pronunciation Software Ltd. at www.eyespeakenglish.com, *Pronunciation Software* by Executive Language Training, www.eltlearn.com, among others. In particular, accent reduction software has become very popular, related to English-speaking countries naturalization and employment issues (e.g., in call centers, where intelligible pronunciation and perfect auditory comprehension are indispensable). Examples of accent reduction systems are *Accent Improvement Software* at www.englishtalkshop.com, *Voice and Accent* by Let's Talk Institute Pvt Ltd. at www.letstalkpodcast.com, *Master the American Accent* by Language Success Press at www.loseaccent.com.

The biggest issue in CAPT application design is to effectively implement the process of learner-system interaction for the program to be able to identify the learner's pronunciation errors and provide a necessary feedback. In other words, the system should operate similar to a human ESL teacher via the basic steps in teaching phonetic phenomena as follows.

1. Explanation: the teacher describes what position the articulatory organs must take and how they must move in order to produce the target sound or sound combination.
2. Imitation: the learner listens to words which contain the target sound and repeat them;

3. Adjustment: the teacher corrects the learner's errors while s/he is imitating the sound/s until its/their production is acceptable;
4. Recognition: the learner listens to input and discriminate the words with the target sound and the words without it.

The problem of human-CAPT system interaction is related to another task of Computer Science called Automatic Speech Recognition (ASR). ASR is a highly complex computational problem and much research effort has been devoted to it; the interested reader may consult the latest ASR advances in [5] and [21]. There have been a number of good efforts to apply ASR results in CAPT systems perusing the two-sided objective, i.e., phonemic recognition of the learner's speech and overall pronunciation assessment or individual error detection [7, 19]; the results obtained at this step are used by the system to generate corrective instructions to the learner.

For pronunciation assessment (evaluation of overall similarity to English speech), the following models have been used: hidden Markov models to calculate the score termed "goodness of pronunciation", GOP [32], Bayesian probabilistic scheme to compute intelligibility levels of students [27], Support Vector Machine to estimate the pronunciation quality score [9], auditory periphery models [12], and their combinations.

In combination with HMM, other strategies have been implemented to detect, or localize individual errors: dynamic time warping (DTW) technique [20], error rules of several types based on articulatory, receptive, and orthographic difficulties [16], Linear Discriminant Analysis [26], phonological rules derived from L1/English contrastive phonologic analysis made on the Cantonese-English language pair [15], error pattern definitions [29], etc.

Though much work has been done in the CAPT field, the challenge of creating an efficient interactive CAPT system still remains. Basically, there are two approaches to pronunciation teaching: the so-called universal approach when training is offered to learners with any L1 without taking it into account, and the L1-specific approach which make error detection more accurate due to mispronunciation prediction based on contrastive phonological analysis. We believe that while existing implementation of ASR techniques does not provide a high quality pronunciation assessment and consequently relevant individualized feedback to the learner, it is more effective to employ L1-oriented approach in tutor systems. To this end, we have made a comparative analysis of American English and Mexican Spanish sounds and based of it we defined some error patterns to be implemented in the error detection module of a CAPT system. We also illustrate our approach with some examples of L1-oriented presentation and explanation of English sounds and phonetic exercises designed in a way that prevents pronunciation errors in the learner's speech.

2.2 ESL Pronunciation for Spanish Speakers

There are scarce resources for Spanish learners of English pronunciation. The fullest courses are *English Phonetics and Phonology for Spanish Speakers* [18] and *A Course in English Phonetics for Spanish Speakers* [8], but they teach British English to Castilian Spanish speakers. Such books like *Teaching English Sounds to Spanish Speakers* [25], *English Pronunciation for Spanish Speakers: Vowels* [3], *English*

Pronunciation for Spanish Speakers: Consonants [4] teach American English, but are limited to some aspects of pronunciation and do not consider Mexican Spanish peculiarities. The approach of all the above mentioned courses is teaching phonemes and very few allophones, if any. This work addresses the lack of teaching resources for American English–Mexican Spanish language pair.

3 Overview of CAPT System Architecture

The basic architecture of a CAPT system includes four principal modules shown in Figure 1. The modules of the system interact with the human learner through interface.

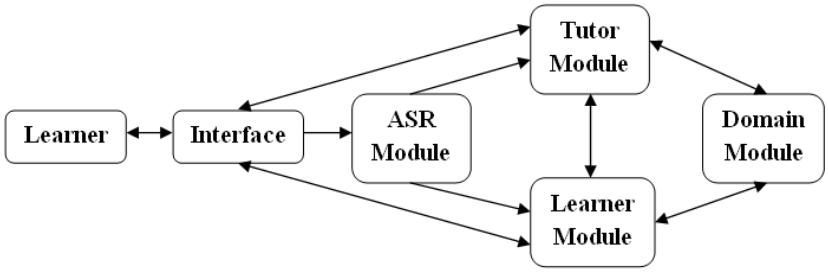

Fig. 1. Basic architecture of a CAPT system.

The tutor module simulates the English teacher; its functions are as follows:

- determine the level of the user (Mexican Spanish-speaking learner of English pronunciation);
- choose a particular training unit according to this learner's prior history stored in the learner's module as data introduced previously via the learner's personal account in the system;
- present the sound or group of sounds corresponding to the chosen training unit and explain its articulation using comparison and analogy with similar sounds in Mexican Spanish;
- perform the training stage supplying the learner with training exercises, determining her errors, generating necessary feedback and selecting appropriate corrective drills;
- evaluate the learner's performance;
- store the learner's scores and error history in the learner's module.

The learner module models the human learner of English; it contains the learner's data base which holds the following information on the learner's prior history:

- training units studied;
- scores obtained;

- errors detected during the stage of articulation training and the auditory comprehension stage.

The domain module contains the knowledge base consisting of two main parts:

- patterns of articulation and pronunciation and auditory perception errors typical for MS speakers as well as individual error samples;
- presentation and explanations of sounds, exercises for training articulation and auditory comprehension.

4 Comparison of American English and Mexican Spanish Vowel Sounds

The basic principle for developing a CAPT system according to the approach presented in this paper is pronunciation error prevention. In this work, we focus on errors in pronouncing AE (American English) sounds. By sounds we mean most frequently met allophones of AE phonemes.

Errors in sound generation may appear for two reasons. Firstly, a target sound may be absent in the source language; secondly, a target sound may exist in the source language but differ to some degree from its counterpart in the source language. In both cases, the learner will tend to substitute the AE sounds with the most similar sounds of her first language thus producing an accent in her speech.

A comparative analysis of AE and MS sound system allows us to predict what MS sounds may be used to substitute the AE sounds and design error patterns for the error detection module of the CAPT system as well as specific strategies for explaining and practicing each AE sound in the most effective way via the system's tutor module.

We have made our best attempt to present almost all allophones of AE and MS phonemes in a way that helps the teacher to predict errors in the AE learner's sound generation and offer her relevant explanation and corrective exercises.

It is not an easy task to compare the sounds of two languages due to the fact that phoneticians have different views on the inventory of phonemes and their allophones as well as on their definitions. Also, although much work has been done in AE phonetic studies, the same cannot be said about Mexican Spanish phonology and phonetics. An additional difficulty is the task itself. As mentioned previously, in order to produce a fluent and least accented AE speech on the one hand and to comprehend what is said in AE on the other hand, the learner should master well all basic allophones (sounds), at the same time relating them to the corresponding phonemes, since they are the basis for the English orthographic system.

Therefore, in the following four subsections we represented AE and MS vowel and diphthong phonemes together with their allophones as well as phonetic features of all included sounds using IPA phonetic alphabet and narrow transcription. The description of sounds relies on the works in [1, 6, 17, 22, 23, 30].

The sounds are described using the following pattern. First, we indicate if a given sound is American English (AE) or Mexican Spanish (MS). Then, the phonetic descriptors, or features, are listed. The phoneme sign is given in forward slashes, and

then a example word is presented. After that, the basic allophones of the sound are given: additional phonetic feature/s distinguishing this allophone is/are specified, the allophone symbol is given in brackets followed by an example word (or words) in which this allophone is used; lastly, we explain in what contexts and under what condition this allophone is produced. Besides, every example word is transcribed; its narrow transcription is given in brackets.

It can be noticed in Sections 4.1–4, that we have adopted a simplistic approach to MS phoneme inventory which takes into account only monophthong phonemes. This point of view is used in [22], for example. According to such approach, the sounds viewed by some phoneticians and Spanish teachers as diphthongs are viewed as combinations of respective phonemes. For example, in the word *aire* the first two vowels pronounced as [ai̯] are considered a diphthong in some phonetic literature, while in works of other researchers it is analyzed as a combination of the basic allophone of the /a/ phoneme and the allophone [i̯] of the /i/ phoneme. By now in this work, we do not consider MS diphthongs.

4.1 Front Vowels

1. **MS high-front /i/ as in *ipo* ['ipo]**. Allophones: nasalized [ĩ] as in *instante* [ĩn'stan̪te], *mimo* ['mĩmo]; occurs between a pause and a nasal consonant or between two nasal consonants; palatal semi-consonant [j] as in *pasión* [pa'sjon], occurs in a prenuclear position; palatal semi-vowel [i̯] as in *aire* ['ai̯re].
2. **AE high-front tense unrounded /i/ as in neat [nit¯]**. Allophones: diphthongized [iɪ] as in *flee* [fl͡iɪ], occurs in open and stressed syllables; diphthongized [iə] as in *seal* [siəl], occurs before a liquid; reduced [ə] or [ɪ] as in *revise* [rə'vaɪz] or [rɪ'vaɪz], in unstressed syllables; lengthened [iː] as in *bee* [biː], word-finally; semi-lengthened [iˑ] as in *been* [biˑn], before a voiced consonant; shortened [i] as in *beat* [bit¯], before a voiceless consonant.
3. **AE lower high/front lax unrounded /ɪ/ as in *bit* [bit¯]**. Allophones: reduced [ə] as in *chalice* ['ʧæləs], in unstressed syllables; lengthened [ɪː] as in *carrying* ['kærɪːŋ], in the position where two /ɪ/ sounds belonging to different morphemes meet.
4. **MS mid-front /e/ as in *este* ['este]**. Allophones: nasalized [ẽ] as in *entre* ['ẽn̪tre], *nene* ['nẽne], between a pause and a nasal consonant or between two nasal consonants.
5. **AE mid-front tense unrounded /e/ as in *ate* [et–]**. Allophones: diphthongized [eɪ] as in *take* [teɪk–], in an open syllable; diphthongized and lengthened [eːɪ] as in *say* [seːɪ], word-finally; diphthongized and semi-lengthened [eˑɪ] as in *name* [neˑim] before a voiced consonant; diphthongized and shortened [eɪ] as in *lake* [leɪk–], before a voiceless consonant; changes to [i] or [ɪ] as in *Monday* ['mʌndɪ], in words with "-*day*".
6. **AE lower mid-front lax unrounded /ɛ/ as in *get* [gɛt–]**. Allophones: diphthongized, r-colored and lengthened [ɛːɚ] as in *tear* [t^{h}ɛːə], word-finally before the letter "r"; diphthongized, r-colored and semi-lengthened [ɛˑɚ] as in *scared* ['skɛ̄ˑɚd–], before the letter "r" followed by a voiced consonant; diphthongized, r-colored and shortened [ɛɚ] as in *scarce* [skɛ̄ɚs], before the letter "r" followed by

a voiceless consonant; triphthongized [eɪə] as in *jail* [dʒeɪəl], before /l/; changes to [ɪ] as in *get* [gɪt–], in informal speech.

7. **AE low-front lax unrounded /æ/ as in *bat* [bæt‾]**. Allophones: lengthened [æ] as in *bad* [bæːd‾] before a voiced consonant; shortened [æ] as in *bat* [bæt], before a voiceless consonant.

4.2 Central Vowels

1. **MS low-central /a/ as in *papa* ['papa]**. Allophones: nasalized [ã] as in *ambos* ['ãmbos], *mano* ['mãno], between a pause and a nasal consonant or between two nasal consonants.
2. **AE lower mid-to-back central lax unrounded /ʌ/ as in *above* [ə'bʌv]**. Allophones: changed to [ɛ] as in *such* [sɛtʃ], in informal speech; changed to [ɪ] as in *just* [dʒɪst], in selected words.
3. **AE neutral mid-central lax unstressed unrounded /ə/ as in *above* [ə'bʌv]**. Allophones: changed to [ɪ] as in *telephone* ['telɪfon], in selected words.
4. **AE mid-central r-colored tense /ɝ/ as in perk [pʰɝk‾]**. Allophones: lengthened [ɝː] as in *sir* [sɝː], word-finally; semi-lengthened [ɝˑ] as in *learn* [lɝˑn], before a voiced consonant, shortened [ɝ] as in *thirst* [θɝst–], before a voiceless consonant.
5. **AE mid-central r-colored lax /ɚ/ as in herder ['hɝdɚ]**. Allophones: r-dropped [ə] as in *motherly* ['mʌðəlɪ], before /r/.

4.3 Back Vowels

1. **AE high-back tense rounded close /u/ as in *boot* [but–]**. Allophones: diphthongized [uə] as in *stool* [stuəl], before a liquid; diphthongized [uʊ] as in *do it* ['duʊɪt–], in stressed or open syllables; reduced [ʊ] or [ə] as in *to own* [tʊ'on], *to go* [tə'go], in unstressed syllables; lengthened [uː] as in *blue* [bluː], word-finally; semi-lengthened [uˑ] as in *food* [fuˑd–], before a voiced consonant; shortened [u] as in *loop* [lup–], before a voiceless consonant.
2. **AE high-back lax rounded /ʊ/ as in *book* [bʊk–]**. Allophones: reduced [ʌ] or [ə] as in *would* [wʌd–] or [wəd–], in rapid speech.
3. **MS mid-back /o/ as in *oso* ['oso]**. Allophones: nasalized [õ] as in *hombre* ['õmbre], *mono* ['mõno], between a pause and a nasal consonant or between two nasal consonants.
4. **AE mid-back tense rounded close /o/ as in *owed* [od–]**. Allophones: diphthongized [oʊ] as in *go* [goʊ], in stressed and open syllables; reduced [ə] as in *window* ['wɪndə], in unstressed syllables; diphthongized and lengthened [oːʊ] as in *no* [noːʊ], word-finally; diphthongized and semi-lengthened [oˑʊ] as in *load* [loˑʊd–], before voiced consonants; diphthongized and shortened [oʊ] as in *coat* [kʰoʊt–], before voiceless consonants.
5. **AE low mid-back lax rounded open /ɔ/ as in *bought* [bɔt–]**. Allophones: lengthened [ɔː] as in *law* [lɔː], word-finally; semi-lengthened [ɔˑ] as in *dawn* [dɔˑn], before voiced consonants; shortened [ɔ] as in *thought* [θɔt–], before voiceless consonants; lowered [ɒ] or [ɑ] as in *cot* [kɒt–] or [kɑt–], after a velar consonant.

6. **AE low-back lax unrounded open /ɑ/ as in *pot* [pɑt̚]**. Allophones: rounded [ɒ] as in *got* [gɒt̚], after a velar consonant; fronted [a] as in *not* [nat̚], after an alveolar consonant; fronted and rounded [ɔ] as in *father* ['fɔðɚ], in platform speech.
7. **MS high-back /u/ as in *pupa* ['pupa]**. Allophones: nasalized [ũ] as in *un soto* ['ũn'soto], *mundo* ['mũn̪do], between a pause and a nasal consonant or between two nasal consonants; velar semi-consonant [w] as in *cuatro* ['kwatro], in a prenuclear position; velar semi-vowel [u̯] as in *auto* ['au̯to], in a postnuclear position.

4.4 Diphthongs

1. **AE rising low-front to high-front /aɪ/ as in *kite* [kaɪt̚]**. Allophones: triphthongized [aɪə] as in *I'll* [aɪəl], before /l/; reduced [ə] *I don't know* [ə'dõʔno], in unstressed syllables in informal speech; lengthened [aːɪ] as in *lie* [laːɪ], word-finally; semi-lengthened [aˑɪ] as in *find* [faˑɪnd̚], before a voiced consonant; shortened [aɪ] as in *light* [laɪt̚], before a voiceless consonant; elevated [ɜɪ] as in *ice* [ɜɪs], before a voiceless consonant.
2. **AE rising low-front to high-back /aʊ/ as in *now* [naʊ]**. Allophones: reduced [ʌʊ] as in *house* [hʌʊs], before a voiceless consonant.
3. **AE rising mid-back to high-front /ɔɪ/ as in voice [vɔɪs]**. Allophones: lengthened [ɔːɪ] as in *boy* [bɔːɪ], word-finally; semi-lengthened [ɔˑɪ] as in *noise* [nɔˑɪz], before a voiced consonant; shortened [ɔɪ] as in *exploit* [əks'plɔɪt̚], before a voiceless consonant.

5 Error Patterns for the Error Detection Module of CAPT System

In this section, some basic error patterns on the phoneme level are presented. They are derived theoretically from the results of comparing AE and MS vowel sound system given in Section 4. Certainly, such theoretical approach is not sufficient to identify all possible errors of an MS learner of English. Practical research is necessary to confirm, clarify, adjust or correct the theoretically predicted errors listed in this section. Also, more error patterns may be discovered in an empirical study of English speech produced by MS learners. We plan to do this research as future work.

Basically, all phoneme errors can be classified into three types:

1. Substitution of an AE phoneme by an MS phoneme.
2. Insertion of an MS phoneme in an AE word.
3. Deletion of an AE phoneme.

In the following three subsections, three types of errors are presented, respectively.

There are two main reasons due to which pronunciation errors are made: the first reason is phonetic, that is, a given AE sound does not exist in MS or if it exists, it differs in some way from it; the second reason is orthographic, when the MS reading rules are applied to AE words. For example, *bat* may be read [bat̚] instead of [bæt̚]

because the letter “a” is read as [a] in all contexts in Spanish. In case an MS learner knows the reading rule of “a” in a closed syllable, she may exhibit a phonetic error of substituting [æ] by [e], since the latter is the MS sound closest the [æ].

In Section 5.1 substitution error patterns are shown. We put the comment “due to orthography”, if an error is made for this reason. In case the reason is phonetic, we give no comment. In Section 5.2 insertion errors are listed, they are caused by the influence of MS orthographic patterns and reading rules. Section 5.3 speaks about deletion errors.

5.1 Substitution

The substitution errors are presented in Table 1.

Table 1. Substitution errors.

AE vowel sounds	Substitution by MS vowel sounds
High-front tense unrounded /i/ as in *neat* [nit‾]	High-front /i/ as in *ipo* ['ipo]
Lower high-front lax unrounded /ɪ/ as in *bit* [bit‾]	
Mid-front tense unrounded /e/ as in *ate* [et–]	Mid-front /e/ as in *este* ['este]
Lower mid-front lax unrounded /ɛ/ as in *get* [gɛt–]	
Low-front lax unrounded /æ/ as in *bat* [bæt‾]	
Mid-central r-colored tense /ɝ/ as in *perk* [pʰɝk‾]	
Mid-central r-colored lax /ɚ/ as in *herder* ['hɝdɚ]	
Neutral mid-central lax unstressed unrounded /ə/ as in *above* [ə'bʌv]	
Mid-front tense unrounded /e/ as in *ate* [et–], due to orthography	Low-central /a/ as in *papa* ['papa]
Low-front lax unrounded /æ/ as in *bat* [bæt‾], due to orthography.	
Lower mid-to-back central lax unrounded /ʌ/ as in *above* [ə'bʌv]	
High-back tense rounded close /u/ as in *boot* [but–]	High-back /u/ as in *pupa* ['pupa].
High-back lax rounded /ʊ/ as in *book* [bʊk–]	
High-back tense rounded close /u/ as in *boot* [but–], due to orthography	Mid-back /o/ as in *oso* ['oso]
High-back lax rounded /ʊ/ as in *book* [bʊk–], due to orthography	
Mid-back tense rounded close /o/ as in *owed* [od–]	
Lower mid-to-back central lax unrounded /ʌ/ as in *above*	

AE vowel sounds	Substitution by MS vowel sounds
[əˈbʌv], due to orthography	
Low mid-back lax rounded open /ɔ/ as in *bought* [bɔt–]	
Low-back lax unrounded open /ɑ/ as in *pot* [pɑt¯]	
Rising low-front to high-back /aʊ/ as in *now* [naʊ], the nucleus /a/ is substituted by MS /o/ due to orthography	
Rising low-front to high-front /aɪ/ as in *kite* [kaɪt¯]	Combination of /a/ and /i/
Rising low-front to high-back /aʊ/ as in *now* [naʊ]	Combination of /a/ and /u/
Rising mid-back to high-front /ɔɪ/ as in *voice* [vɔɪs]	Combination of /o/ and /i/

5.2 Insertion

In Table 2 we present the insertion errors.

Table 2. Insertion errors

MS insertion	In AE words like
/r/ after /ɝ/	*perk* [pʰɝk¯]
/r/ after /ɝ/	*herder* [ˈhɝdɚ]
/u/ after /ɔ/	*bought* [bɔt–]
/a/ after /i/	*neat* [nit¯]
/e/ after /t/	*ate* [et–]

5.3 Deletion

Compared to other types of errors, that is, substitution and insertion of phonemes, deletion of a vowel phoneme or its component is not an error very frequently exhibited by MS learners. The phenomenon of phoneme deletion is more typical for consonant sounds, especially in word final positions. Deletion of a vowel sound occurs mainly for orthographic reasons. For example, in the word *note* [noʊt¯], the second component of the diphthongized allophone of the phoneme /o/ may be deleted because the letter "o" is read as [o] in all context in Spanish.

6 Error Patterns in the Error Detection Module of CAPT System

One of the objectives of learning a second language is to develop speech production and speech recognition abilities. English learners are expected to understand AE speech as well as to realize their communicative intent generating speech in a way that is less accented and intelligible to native speakers. In view of this task, error detection and correction are seen as a very important part of language learning.

In the CAPT system architecture described in Section 3, the learner's speech is processed by the Automatic Speech Recognition (ASR) module, and the error identification function is performed by the Error Detection module.

Automatic error detection at the level of individual sounds is a complex computational task; it remains a challenge in CAPT system development. Compared to human judgment, automatic erroneous sound detection in CAPT systems is not al all satisfactory [26]. Error detection rate can be improved if the error detection module is fed with error patterns to be used as guidelines for predicting errors in learner's speech.

Modern CAPT systems commonly use Hidden Markov Models for error detection. Let us consider the process of isolated word recognition since pronunciation training begins with mastering AE sounds in individual words. The process includes two major stages. Firstly, the ASR system is trained using a vocabulary of pronounced words mapped to their transcriptions which constitutes a phonetic database. Pronounced words are represented as speech vectors. Secondly, the system is exposed to unknown words (speech vectors) and generates their transcription. At the training stage, a HMM is trained for each word using a set of examples of that word. At the recognition stage, when an unknown word is presented to the system, it calculates the likelihood of each model generating that word and the most likely model is chosen. To build the system, Hidden Markov Models Toolkit [33] can be used.

In the words used at the training stage, each vowel sound prone to error can be aligned to a list of errors with their respective probabilities. The probabilities can be estimated in an empirical study of English speech produced by MS learners as a part of future research. Here we give the probability values as supposed by us based on theoretic research. These values may be used as a starting guess for initiating HMM models.

Table 3. Vowel pronunciation errors in the word *road*

Correct	Incorrect		
[roʊd⁻]	Transcription	Probability	Reason
	[road⁻]	0.6	Orthographic
	[roud⁻]	0.35	Substitution of /ʊ/ with /u/
	[rod⁻]	0.05	Deletion of /ʊ/

Consider the word *road* [roʊd⁻] (Table 3). Two transcriptions will be stored in the phonetic database: the correct transcription and the transcription including possible erroneous sounds annotated with their probabilities. In case the word exposed to the system differ significantly from the correct version based on a pre-defined threshold,

the system will take into account error pattern probabilities in order to identify the concrete error. We consider only errors in vowel sound pronunciation. In a complete model, all sounds, vowels and consonants, are considered.

7 Examples of Error-Preventive AE Sound Training for the Tutor Module of CAPT System

In this section we give two examples of teaching AE sounds to MS speakers taking into account the comparative analysis of AE and MS sounds presented in Section 4. The first example is described in more detail, and the second one is presented in a more concise way since the training stages in the second example are the same as in the first one.

The phoneme teaching is realized in the following stages:

1. AE phoneme presentation and explanation of its articulation in comparison with similar MS sound/s.
2. Training of the AE phoneme first using MS words with similar sound/s, then AE words of increasing complexity.
3. Training of auditory recognition of the AE phoneme first using minimal pairs, then words of increasing complexity, word combinations and phrases depending on the learner's level (elementary. intermediate, advanced).

In order to prevent errors in sound generation, the MS learner needs to understand the differences between the AE and MS sounds on the articulatory level as well as on the auditory level. Besides, she has to learn to relate the articulatory movements with the auditory effects produced by them, because this ability is fundamental for adequate speech recognition.

Taking these objectives into account, we suggest introducing, explaining and practicing the AE sounds not in minimal pairs but in so-called **minimal triplets**, adding to each minimal AE word pair a Spanish word containing a similar Spanish sound. Such MS sound may substitute the target AE sound/s and produce misunderstanding of English speech. For this reason, the learner should understand the difference in articulation and acoustics of AE and MS sounds in order to be able to produce less accented speech and avoid misunderstanding in oral recognition. Minimal triplets can be used at the elementary level of AE pronunciation training. When the AE sounds are pronounced sufficiently well by the learner, only AE words and phrases will be used in further stages.

As an example, we chose two AE phonemes: /u/ and /ʊ/. The phonemes /u/ as in *boot* [but–] and /ʊ/ as in *book* [bʊk–] are similar to MS /u/ as in *pupa*. For such reason these AE phonemes are substituted by the MS /u/, see substitution error patterns in Table 1. The phoneme /u/ as in *boot* [but–] is high-back tense rounded close, the phoneme /ʊ/ as in *book* [bʊk–] is high-back lax rounded, while the MS phoneme /u/ as in *pupa* is high-back.

Due to the fact that the phoneme pair /u/ – /ʊ/ is absent in MS, words having these phonemes are often confused by MS speakers which decreases their auditory

recognition capacity. In speech generation, both /u/ and /ʊ/ of American English may be substituted by MS /u/, as mentioned before.

At Stage 1, both AE phonemes are presented and explained in contrast with the MS /u/. Their similarities and differences should be clarified in detail accompanied by contrastive examples of minimal triplets. For example, the words listed below can be used. Remember, that two of them are English and the last word is Spanish.

curso	*cool*	*cook*
julio	*who*	*hook*
tubo	*tool*	*took*
gusto	*goose*	*good*
luz	*loom*	*look*
nunca	*noon*	*nook*

The articulation of all three sounds can be illustrated by diagrams showing the movements of the speech organs. The main difference between AE and MS articulation is that the AE phonemes are rounded and the corresponding MS phoneme is not.

At Stage 2, we suggest first to train the AE /u/ phoneme since is it closer to the corresponding MS phoneme. Here we suggest a method that can be called **building** of the AE /u/ on the foundation of its MS counterpart.

The learner starts from her familiar sound /u/ in a word like *pupa* and is told that this sound will be used as a basis for building the corresponding AE sound. Since the sounds are different, the learner has to change the articulation of /u/. For this purpose, the following language therapy exercise may be suggested: the learner is invited to pronounce the MS word *pupa* paying special attention to the /u/ sound and prolonging it, at the same time stretching her lips a little with her hands as in a smile thus avoiding the lip forward protraction typical for MS. The learner should listen carefully to the auditory difference which the lip stretching produces. At the same time, the learner is invited to observe her lip position and movement in a mirror and to form a narrow rounded mouth opening while slightly stretching the lips. The mirror also adds visual control for a more effective acquisition. When the learner understands the difference and is able to produce "the English version" of the Spanish /u/, more words are given for articulation training and auditory recognition exercises.

The phoneme /ʊ/ is built on the foundation of the AE /u/ when the latter is generated adequately. At this step, AE **minimal pairs** may be used, since the learner is supposed to have overcome her natural tendency to substitute the AE /u/ sound with the MS /u/. As an additional practice, the learner may be exposed to pairs of words in which the first word is an AE word containing the sound /ʊ/, and the second one is a similar MS word with the /u/ sound. This exercise will give another opportunity to the learner to reinforce her phonetic awareness of the difference between the AE /ʊ/ and the MS /u/ and to improve her pronunciation and recognition skills with respect to this sound pair.

Another example is teaching the AE phonemes /æ/ and /e/ in contrast with the similar MS phoneme /e/. The AE /æ/ as in *bat* [bæt⁻] is low-front lax unrounded, the AE /e/ as in *ate* [et–] is mid-front tense unrounded, and the MS similar phoneme /e/ as in *este* ['este] is mid-front. Since the stages of teaching the AE /æ/ and /e/ are the same as of teaching any AE sounds, we do no explain each and every detail of the teaching

and acquisition process. Instead, we offer an example of minimal triplets and present a language therapy exercise for building /æ/ on the foundation of the MS /e/, then the AE /e/ is built on the foundation of /æ/. The minimal triplets are as follows.

mes	*mass*	*mess*
necio	*nag*	*neck*
beca	*back*	*beg*
seja	*sad*	*set*
texto	*tan*	*text*
queso	*cap*	*keg*

It may seem strange that we propose first to build /æ/ on the basis of the MS /e/, but not the AE /e/ on the basis of the same MS sound. The reason for this choice is that the difference between the AE /e/ and the MS /e/ is more subtle than the difference between /æ/ and the MS /e/. For a phonetically inexperienced learner it will be more difficult to perceive and produce such difference. Therefore, we think it is better to work on /æ/ it the beginning due to a bigger contrast which is easier for the learner to recognize and produce.

To build the /æ/ sound on the basis of the MS /e/, we suggest the following language therapy exercise. First, the learner is invited to pronounce the MS word *mes* slowly, prolonging the sound /e/. Then, while pronouncing the sound /e/ in *mes*, open the mouth wider, stretch the lips a little as in a smile. At this point it is important to avoid generating /a/ instead of /æ/. The learner is explained that opening her mouth will most probably produce the MS /a/ which should not be done. To prevent /a/ production, the learner is asked to hold the middle part of the tongue in a low position with the help of a spoon pressing the tongue slightly (without applying too much force to prevent undesirable physiological reaction), because the tongue elevation causes /a/ production. If the mouth is opened wider than in the MS /e/, the lips are slightly stretched and the middle part of the tongue is in its lower position, then the AE sound /a/ is generated.

The AE /e/ is trained on the basis of /æ/, asking the learner to open her mouth less than for the sound /æ/, but keeping the tongue in the same position as for /æ/ to avoid the MS /e/ generation.

8 Conclusions and Future Work

In this paper, we presented a detailed comparative analysis of American English and Mexican Spanish sound systems on the level of both phonemes and allophones. The results of this analysis can be used as a basis for creating the linguistic content of error prevention oriented CAPT system. Since error detection is a very difficult task of automatic speech recognition in intelligent tutor systems, its performance can be improved if a first language oriented approach in teaching English pronunciation is adopted. In our work, we considered Mexican Spanish and presented examples of how teaching articulation and auditory comprehension can be enhanced when typical error patterns are known in advance. In future, we plan to implement the results of our

phonetic analysis in designing a robust CAPT system for Mexican Spanish speakers and conduct experiments to compare such system with existing generally oriented tutor systems.

References

1. Avery, P., Ehrlich, S.: Teaching American English Pronunciation. Oxford University Press, England (1992)
2. Burbules, N. C.: Ubiquitous Learning and the Future of Teaching. Encounters on Education, vol. 13, pp. 3–14 (2012)
3. Dale, P.: English Pronunciation for Spanish Speakers: Vowels. Prentice Hall Regents, NJ (1985)
4. Dale, P., Poms, L.: English Pronunciation for Spanish Speakers: Consonants. Prentice Hall Regents, NJ (1986)
5. DeMori, R., Suen, C. Y.: New Systems and Architectures for Automatic Speech Recognition and Synthesis. Springer-Verlag NY Inc. (2012)
6. Edwards, H. T.: Applied Phonetics: the Sounds of American English. Singular Pub. Group, San Diego, CA (1997)
7. Eskenazi, M.: An overview of spoken language technology for education. Speech Communication, vol. 51(10), pp.832–844 (2009)
8. Finch, D. F., Ortiz Lira H.: A Course in English Phonetics for Spanish Speakers. Heinemann Educational Books Ltd, London (1982)
9. Ge, F., Pan, F., Liu, C., Dong, B., Chan, S. D., Zhu, X., & Yan, Y.: An svm-based mandarin pronunciation quality assessment system. In: The Sixth International Symposium on Neural Networks, pp.255–265. Springer Berlin Heidelberg (2009).
10. Ipek, H.: Comparing and Contrasting First and Second Language Acquisition: Implications for Language Teachers. English Language Teaching, vol. *2*(2), pp.155–163 (2009)
11. Khan, B. H.: A Comprehensive E-Learning Model. Journal of e-Learning and Knowledge Society, vol. 1, pp.33–43 (2005).
12. Koniaris, C., Salvi, G., Engwall, O.: On mispronunciation analysis of individual foreign speakers using auditory periphery models. Speech Communication (2013)
13. Levy, M., Stockwell, G.: CALL Dimensions: Options and Issues in Computer-Assisted Language Learning. Lawrence Erlbaum Associates, Inc., NJ (2006)
14. Liakin, D.: Mobile-Assisted Learning in the Second Language Classroom. International Journal of Information Technology & Computer Science, vol. 8(2), pp.58–65 (2013).
15. Meng, H., Lo, W. K., Harrison, A. M., Lee, P., Wong, K. H., Leung, W. K., Meng, F.: Development of automatic speech recognition and synthesis technologies to support Chinese learners of English: The CUHK experience. Proceedings of APSIPA (2010)
16. Menzel, W., Herron, D., Bonaventura, P., Morton, R.: Automatic detection and correction of non-native English pronunciations. Proceedings of INSTILL, pp.49–56 (2000).

17. Moreno de Alba, J. G.: El español en América. Fondo de cultura económica, México (2001)
18. Mott, B. L.: English Phonetics and Phonology for Spanish Speakers. Edicions Universitat de Barcelona, Barcelona (2005).
19. Neri, A., Cucchiarini, C., Strik, W.: Automatic speech recognition for second language learning: how and why it actually works. In: Proceedings of ICPhS, pp.1157–1160 (2003).
20. Nouza, J.: Training speech through visual feedback patterns. In: Proceedings of ICSLP (1998)
21. Pieraccini, R.: The Voice in the Machine. MIT (2002)
22. Pineda, L.A., Castellanos, H., Cuétara, J., Galescu, L., Juárez, J., Llisterri, L., Pérez, P., Villaseñor, L.: The Corpus DIMEx100: Transcription and Evaluation. Language Resources and Evaluation, *vol.* 44(4), pp.347–370 (2010)
23. Quilis, A.: El comentario fonológico y fonético de textos: teoría y práctica. 3a edición. Arco/Libros, S.L., Madrid (1997)
24. Roberts, T. A.: Articulation Accuracy and Vocabulary Size Contributions to Phonemic Awareness and Word Reading in English Language Learners. Journal of Educational Psychology, vol. 97(4), pp.601–616, (2005)
25. Schneider, L. C.: Teaching English Sounds to Spanish Speakers. Allied Educational Council (1971)
26. Strik, H., Truong, K., de Wet, F., Cucchiarini, C.: Comparing different approaches for automatic pronunciation error detection. Speech Communication, vol. 51(10), pp. 845–852, doi: 10.1016/j.specom.2009.05.007 (2009)
27. Tsubota, Y., Kawahara, T., Dantsuji, M.: Practical use of English pronunciation system for Japanese students in the CALL classroom. In: *Proceedings of ICSLP*, pp.849–852 (2004).
28. Vihman, M. M.: Phonological development: The origins of language in the child. Blackwell, Oxford (1996)
29. Wang, Y. B., Lee, L. S.: Improved approaches of modeling and detecting error patterns with empirical analysis for computer-aided pronunciation training. In Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp.5049–5052 (2012).
30. Whitley, M.S.: Spanish-English Contrasts: A Course in Spanish linguistics. Georgetown University Press, Washington, D.C. (1986)
31. Willson, V. L., Rupley, W. H.: A structural equation model for reading comprehension based on background, phonemic and strategy knowledge. Scientific Studies of Reading, vol. 1, pp.45–63 (1997)
32. Witt, S., Young, S.: Computer-aided pronunciation teaching based on automatic speech recognition. In: Jager, S., Nerbonne, J. A., van Essen, A. J. (eds.) Language teaching and language technology, pp. 25–35, Swets & Zeitlinger, Lisse (1998)
33. Young, S., Jansen, J., Odell, J., Ollason, D., Woodland, P.: The HTK Book. Cambridge University (1996)