Error Patterns for Automatic Error Detection in Computer Assisted Pronunciation Training Systems

Olga Kolesnikova

Superior School of Computer Sciences, Instituto Politécnico Nacional,
Mexico City, Mexico

kolesolga@gmail.com

Abstract. This paper presents error patterns built on the basis of our comparative analysis of American English and Mexican Spanish phonemes and allophones which can be applied in designing the error detection module of a Computer Assisted Pronunciation Training (CAPT) System for teaching American English pronunciation to Mexican Spanish speakers. Error identification is important for an adequate choice of correcting techniques which improves English pronunciation acquisition and helps learners to develop less accented speech. Since automatic individual error detection remains a highly complex computational task, error patterns can enhance the system performance and improve its precision. To the best of our knowledge, error patterns in American English speech generated by Mexican Spanish speakers has not been defined in previous work which was done mainly for Castilian-originated standard Spanish.

Keywords: Error patterns, pronunciation, Mexican Spanish.

1 Introduction

In second language (L2) learning, it is very important to acquire reasonably correct pronunciation. We consider reasonably correct pronunciation because, speaking in terms of general public, it is very hard to develop perfect, native-like L2 pronunciation. Usually, some accent is acceptable whenever the speech of an L2 learner is comprehensible to L2 native speakers.

Correct pronunciation is important not only for L2 learners to be understood adequately, but also for them to understand L2 native speakers. It is a typical problem in L2 learning process that a learner can speak and read, but it becomes a real pain in the neck when it comes to listening comprehension of real-life everyday speech which is usually characterized by high speech, sound reduction, and phonetic variation. Here, the acquisition of correct pronunciation can help since the articulatory and auditory systems are interconnected. A learner is hardly able to recognize a sound which she has never produced due to its absence in her first language (L1). So if a learner has acquired the correct articulation of an L2 sound in its isolate position and in combinations, and has devoted sufficient time to practicing its production, she will be able to recognize it in fluent L2 speech.

In the beginning, we mentioned that it is generally acceptable if an L2 learner develops a reasonably correct pronunciation. However, in some cases, for some activities and occupations, less or even non-accented speech is a requirement. An example of such jobs is operators in call centers. Here, an L2 learner will need more pronunciation training than general language teaching courses can provide, and would look for a specialized course, classes, or software which presents this phonological and phonetic aspect of L2 in more detail.

Nowadays, Computer Assisted Language Learning (CALL) is recognized as a beneficial tool for both L2 teachers and learners. Accessibility in practically all everyday situations, flexibility, adaptability and personalization make CALL systems an excellent instrument in any kind of learning: group and individual, formal and informal, stationary and mobile, in and outside classroom [2, 11, 12, 13].

Many CALL applications are designed to facilitate L2 acquisition in all language aspects: pronunciation, words and their usage, grammar, pragmatics. But there are tutor systems created for Computer Assisted Pronunciation Training (CAPT). A variety of CAPT commercial software can be found online: *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.

Responding to the user's particular need of reducing L2 accent necessary to resolve naturalization and employment issues in English-speaking countries, specialized accent improvement systems have recently been produced. Some examples 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.

In this paper, we will look at a particular aspect of CAPT systems, namely, their capability of recognizing, or localizing, errors in the learner's speech implemented in the error detection module of the system. Error identification is important for generating an appropriate feedback and corrective exercises to the learner by means of the tutor module of the same system with the purpose of improving the learner's pronunciation and listening comprehension.

Since automatic individual error detection remains a highly complex computational task, error patterns, or error rules, can enhance the system performance and improve its precision. In this work, we define error patterns typical for American English (AE) speech generated by Mexican Spanish (MS) native speakers. The defined error patterns are based on our comparative analysis of AE and MS phonemes and allophones. To the best of our knowledge, such error patterns have not been defined in previous work which was done mainly for Castilian-originated standard Spanish.

The rest of the paper is organized as follows. Section 2 considers the impact of error identification and adequate treatment in the process of L2 acquisition. Section 3 presents the basic architecture of a Computer Assisted Pronunciation Training system and its modules. Section 4 surveys the implementation of Automatic Speech Recognition (ASR) techniques in CAPT systems and briefly describes four essential ASR steps. Section 5 considers two approaches for detecting errors in CAPT systems: general pronunciation assessment and individual error detection. In Section 6 we present error patterns determined on the basis of our comparative analysis of AE and MS phonemes and allophones, and Section 7 outlines conclusions and future work.

2 Errors in the Process of Second Language Acquisition

Traditional language courses teach pronunciation and auditory recognition of second language phonemes using four basic steps listed as follows, each step is termed twice: first, using general pedagogic terminology, and second, referring to the processes in an intelligent tutor system designed to implement these steps.

At Step 1, which may be called explanation (input), 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. At Step 2, imitation (output), the learner listens to words with the target sound and repeats them. At Step 3, adjustment (feedback), the teacher identifies, explains, and corrects errors of the learner with relevant exercises until production of the target sound is appropriate depending on the orientation of the course and the learner's level. At Step 4, recognition (assessment), the learner listens to input and discriminates words with the target sound and words without it.

Special attention is paid to correcting the learner's errors at Step 3. Making first articulatory attempts in L2, learners almost always make errors, especially if the phoneme they are practicing at the moment is not present in their L1. In fact, committing and correcting errors is a common aspect of the language learning process. Therefore, it is important for a human teacher or a computer tutoring system to identify errors in the learners' speech, to explain the causes of such error and to offer adequate corrective exercises.

Speaking about intelligent tutor systems, we should mention that their error detecting capacity remains an open question in computer science. Notwithstanding the impressive technological advance we are witnessing now, CAPT systems still require further improvement. The system's capacity to detect errors in the speech of the learner and to offer a relevant feedback —activities performed at Step 3 of the teaching/learning process— is an issue of ongoing research.

In this paper, we focus on this important challenge and address it by defining error patterns to be implemented in the error detection module of a CAPT system to teach American English (AE) pronunciation to native speakers of Mexican Spanish (MS). On the other hand, the same error patterns can be used in the tutor module of a CAPT system in the manner which prevents possible errors and develops new sound generation and auditory recognition skills on the foundation of similar L1 sounds. We believe that such approach will make the process of pronunciation acquisition conscious at all stages (important mostly for adult learners) and free of stress and awkward feelings caused by the fact that learners face the necessity of generating sounds completely alien to them.

3 CAPT Systems

As it was mentioned in the Introduction, Computer Assisted Language Learning (CALL) in general and Computer Assisted Pronunciation Training (CAPT) in particular expand to a great degree opportunities for learners to study independently in a non-judgmental context at their own pace and their preferred location, to view and/or review any part of the materials, to enjoy a variety of practice and to get an individualized corrective feedback.

In this work we are interested in CAPT applications oriented at teaching English as a second language. Although the advantages of such applications are beyond doubt, there are some issues which have not been efficiently resolved yet [9]. These problems include a lack of pedagogical foundation, an emphasis on practicing pronunciation of individual words outside of their context and not in connected speech, insufficient training of suprasegmental features of pronunciation (stress, tone, word juncture) as well as a poor quality of feedback.

The main reason why computer feedback sometimes fails to provide meaningful and relevant assistance to learners is a high complexity of the task which a system has to solve: it must be able to process the learner's speech, identify the pronounced words/phrases and detect errors in them. The area of computer science which deals with these and similar tasks is called Automatic Speech Recognition (ASR), and automatic error detection is a part of ASR.

Quite a lot of research effort has been devoted to solve speech recognition problems; the interested reader may consult some recent ASR advances in [4, 17]. Also, there have been a number of attempts to apply ASR results in CAPT systems perusing the two-sided objective: phonemic recognition of the learner's speech and overall pronunciation assessment or individual error detection [6, 16]; the results obtained at this step are used by the system to generate corrective instructions to the learner. In spite of a progress in improving the quality of computer produced feedback, it is still not as satisfactory as it is expected to be. This work is another attempt to deal with the issue of automatic error detection in CAPT systems in order to improve the precision of the CAPT system feedback.

Usually, in the architecture of CAPT systems, the function of error detection is represented by a separate module. Now we will describe the overall design of such system, explain the functions of each basic module of the system and in Section 5 devoted to error detection we will discuss state of the art techniques implemented in the error detection module of modern CAPT systems.

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.

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 in our work); choose a particular training unit according to the 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.

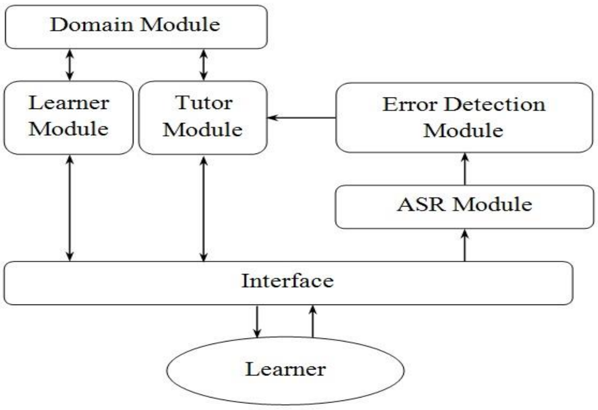


Fig. 1. Basic architecture of a CAPT system

The domain module contains the knowledge base consisting of two main parts: patterns of articulation and pronunciation and auditory perception error patterns characteristic of MS speakers as well as individual error samples; presentation and explanations of sounds, exercises for training articulation and auditory comprehension.

The ASR module is responsible for recognition of the learner's speech.

The error detection module processes the output of the ASR module and identifies pronunciation errors.

4 Automatic Speech Recognition in CAPT Systems

The basic goal of Automatic Speech Recognition (ASR) is to take an acoustic waveform as input and produce a string of words as output. Such analysis involves segmentation of fluent speech into units called phones.

A phone is a speech segment with distinct physical and perceptual (articulatory, acoustic, auditory) features which is a basic unit of phonetic speech analysis, in other words, as we view it, it is a speech sound. The term *phone* is preferred to *sound* in ASR literature although in our opinion both words denote the same entity.

To represent phones, a phonetic alphabet is used. There exist a number of phonetic alphabets, and in this work we use the IPA (International Phonetic Association) phonetic alphabet; see *Handbook of the International Phonetic Association* [8]. The official website of IPA is at http://www.langsci.ucl.ac.uk.

Another important concept used in ASR is a phoneme which is a "contrastive segment" of speech. The word *contrastive* means that one segment (a phoneme) contrasts with other segments "to make a change in meaning" [3:p.41]. For example, in the four words *cat* [k^{h}æt], *pat* [p^{h}æt], *rat* [ræt], *chat* [tʃæt] the only "segment of speech" which differs is the one at the beginning of each word, and this difference produces a change in meaning, so this fact identifies in this case four different phonemes: /k/, /p/, /r/, and /t/.

However, each phoneme can be pronounced in various manners, for example, /k/ in *cat* can be pronounced with aspiration as [k^h] or without aspiration as [k] but such variation does not change the meaning of *cat*, so [k^h] and [k] are not different phonemes, but they are different phones. Phone symbols are written in brackets to distinguish them from phonemes. If two or more phones are realizations of the same phoneme, such phones are called allophones. In our example, [k^h] and [k] are two allophones of the phoneme /k/.

Automatic speech recognition is a complex process consisting of several stages. In summary, a speech signal is first processed to be represented in a computer (this part is called signal processing), and then such representation is analyzed with the purpose of determining to which word or words a given signal corresponds (this part is called signal decoding).

Now, in a more detailed overview, the ASR process can be described by four steps which we are going to discuss now. Each step is considered in a separate subsection. Figure 2 presents a diagram of the four ASR steps.

4.1 Step 1: Speech Waveform Segmentation

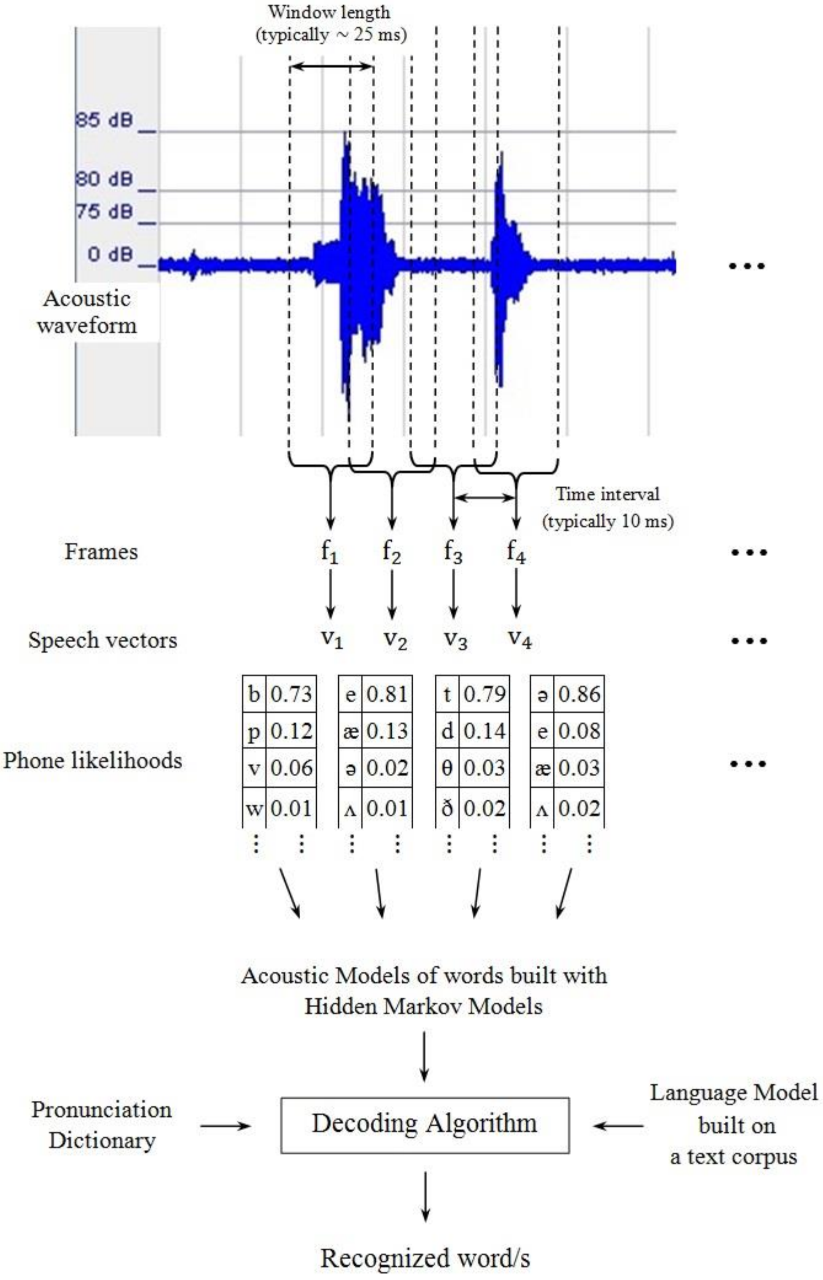
At the first stage, a speech signal —an acoustic waveform— is processed to be represented in a computer system. For speech processing, various methods, analog and digital, are used. A common approach is to view a speech signal as a function of time.

However, there are many factors involved in speech production, and speech characteristics change constantly. But if we cut a speech signal into very small intervals of 5 to 25 ms, speech characteristics can be viewed as constants, and intervals of analysis can be mapped to individual phones (at the next stage of ASR).

So at the stage of signal processing, an acoustic waveform is segmented into small pieces called frames. The length of a frame is called the window length. The latter is a parameter, it can be set depending on what information we want to extract from a signal. Also, at this stage which is called the segmentation stage, slices are made in such a way that there is usually a 50% of overlap between two succeeding frames.

4.2 Step 2: Speech Parametrization

At the second stage, each frame is represented by means of a speech vector, or a spectral feature vector. The purpose of this step is to present the speech waveform under analysis in a compact form and to extract information necessary and sufficient to distinguish one phone (of the inventory, usually about 40 for the English language) from another and filter out acoustic information characteristic of individual speakers.



b	0.73	e	0.81	t	0.79	ə	0.86
p	0.12	æ	0.13	d	0.14	e	0.08
v	0.06	ə	0.02	θ	0.03	æ	0.03
w	0.01	ʌ	0.01	ð	0.02	ʌ	0.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Fig. 2. Automatic speech recognition basic steps

Here we are interested in creating the general acoustic image of a phone "on average", independent of peculiarities of individual pronunciation. This stage of speech processing is called the speech parametrization stage, or the encoding stage.

Various algorithms exist for speech parametrization; the most frequently used of them as well as the simplest one is the procedure based on Mel-frequency cepstral coefficients (MFCCs). The output of this algorithm is a feature vector whose dimensionality is about 40.

4.3 Step 3: Phone Recognition

At the third stage of ASR, speech vectors are mapped to phones or groups of phones; therefore, this stage is called the phone recognition stage. Here, statistical techniques are used such as neural networks or Gaussian models. However, the most common technique used for estimation phone likelihoods is Hidden Markov Models. The underlying idea is to calculate the probabilities for each frame to correspond to each phone in the inventory.

4.4 Step 4: Decoding

The final stage of ASR is called the decoding stage. Now, the probabilities calculated at the previous stage are used to determine to which word in the dictionary a given signal corresponds. At this stage a Viterbi Decoder or a Stack Decoder is used. The input to the decoder is an acoustic model of the utterance to be recognized, a pronunciation dictionary, and a language model. Language model is usually an n-gram model built on a sufficiently large text corpus. The output of the decoder is the recognized word or words.

5 Error Detection in CAPT Systems

In modern CAPT systems, there are two approaches to the learner's pronunciation evaluation and error detection [6]. According to the first approach, the system performs an overall learner's pronunciation evaluation and calculates a measure of **pronunciation assessment**. Within the frame of the second approach, called **individual error detection**, the system detects particular errors of a learner which is a much more difficult issue compared to the first approach due to computational complexity of the ASR task in general and unresolved problems of individual sound recognition in particular, so this issue is still an open question and an area of ongoing research.

5.1 Pronunciation Assessment

Pronunciation assessment is an evaluation of overall impression of the L2 learner's fluent speech. The best known measure is Goodness of Pronunciation (GOP) proposed by Witt [24].

GOP gives a score for each phone of an utterance. To calculate the scores, the GOP algorithm uses the orthographic transcription of the pronounced utterance under analysis and a set of Hidden Markov Models which determine the likelihood $p\left(O^{(q)}\middle|q\right)$ of the acoustic segment $O^{(q)}$ corresponding to each phone q. Then, the quality of pronunciation for any phone p is the duration normalized log of the posterior probability $P\left(p\middle|O^{(p)}\right)$ that the speaker uttered phone p given the corresponding acoustic segment $O^{(p)}$:

$$\mathrm{GOP}(p)=\frac{\left|\log\left(P\left(p\middle|O^{(p)}\right)\right)\right|}{NF(p)}=\frac{\left|\log\left(\frac{p\left(O^{(p)}\middle|p\right)P(p)}{\sum_{q\in Q}p\left(O^{(p)}|q\right)P(q)}\right)\right|}{NF(p)}$$

where Q is the set of all phone models and $NF(p)$ is the number of frames in the acoustic segment $O^{(p)}$. If it is assumed that the likelihood of all phones is the same, i.e., $P(p)=P(q)$, and that the sum $\sum_{q\in Q}p\left(O^{(p)}\middle|q\right)P(q)$ can be approximated by its maximum, the formula becomes

$$\mathrm{GOP}(p)=\frac{\left|\log\left(\frac{p\left(O^{(p)}\middle|p\right)}{\max\limits_{q\in Q}p\left(O^{(p)}|q\right)}\right)\right|}{NF(p)}.$$

5.2 Individual Error Detection

Up to now, attempting to develop a good performance technique for individual error detection, researches have suggested a number of strategies, most representative of which are briefly reviewed in this section.

Weigelt *et al.* [22] used decision trees to discriminate between voiceless fricatives and voiceless plosive using three measures of the waveform. The authors did not apply their results directly to error detection although such application was implied. Later, this method was applied by Truong *et al.* [21] to identify errors in three Dutch sounds /A/, /Y/ and /x/, often pronounced incorrectly by L2-learners of Dutch. The classifiers used acoustic-phonetic features (amplitude, rate of rise, duration) to discriminate correct realizations of these sounds. Truong *et al.* [21] also used classifiers based on Linear Discriminant Analysis (LDA) obtaining good results. Strik *et al.* [20] is another work which experimented with the method in [22] and compared it to other three methods, namely, Goodness of Pronunciation, Linear Discriminant Analysis with acoustic-phonetic features, and Linear Discriminant Analysis with mel-frequency cepstrum coefficients. The analysis was done for the same three Dutch sounds as in [21].

Error detection task was studied for languages other than Dutch. Zhao *et al*. [26] used Support Vector Machines with structural features to identify Chinese pronunciation errors of Japanese learners. Decision tree algorithm was used in the work of Ito *et al*. [10] to identify English pronunciation errors in the speech of Japanese native speakers. The same task was pursued for Korean learners of English in the work of Yoon *et al*. [25] using a combination of confidence scoring at the phone level and landmark-based Support Vector Machines. Menzel *et al*. [14] used the confidence scores provided by a HMM-based speech recognizer to localize English pronunciation error of Italian and German speakers.

It is natural that English as a second language attract attention of researchers who are speakers of other languages. In our work, we are interested in error detection to be implemented in an L2-English pronunciation training system for Mexican Spanish native speakers. For the error detection module of the system we have chosen an approach of error pattern definitions on the basis of comparative analysis of the sound systems of the two languages. Since errors in the patterns have higher probability of occurrence, this fact can improve the overall accuracy of the automatic error detection which by now is not at all satisfactory in state of the art pronunciation training software (see criticism of CAPT systems by Neri *et al*. [16]).

6 Error Patterns for CAPT Systems

Meanwhile the task of individual error detection in any language remains highly complex and not resolved to a satisfactory degree, most state of the art CAPT systems have been designed for pairs of languages, of which one language is usually the native language of a learner, and the other is the second language she is mastering with the help of the CAPT system. In such case, individual error detection is facilitated by the knowledge of typical errors the learner can make.

Typical errors can be encountered theoretically and/or empirically. Theoretical error identification is performed by means of a comparative phonetic analysis of sounds, usually phonemes, and empirical error detection is done based on a study of learner corpora [7].

For example, English learner corpora include recorded interviews, read texts, conversations and other samples of spoken English produced by non-native English speakers. A recognized and oft-used English learner corpus is *Louvain International Database of Spoken English Interlanguage* (LINDSEI, at http://www.uclouvain.be/en-cecl-lindsei.html). This corpus contains oral data produced by advanced learners of English from several mother tongue backgrounds including Bulgarian, Chinese, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish, and Swedish. It includes almost 800,000 words produced by learners, which represents 554 interviews corresponding to more than 130 hours of recording.

In this paper, we define error patters according to the comparison of American English and Mexican Spanish sounds at the level of allophones which takes into account phonetic processes in both languages. In state of the art works on this theme, analyses are made typically at the level of phonemes only. However, a speaking person does not produce phonemes (abstract units with the capability of distinguishing meaning), but allophones, i.e., realizations of phonemes in real speech.

Certainly, there exist a very big number of allophones due to language variability depending on phonetic processes, individual articulatory characteristics of a person, his or her educational level, social status, location, age, etc., and it is not feasible to identify all of such allophones and use them in CAPT systems. However, concerning allophones and their acquisition in the process of L2 learning, most frequently met allophones in standard speech should be selected. For American English, General American accent is considered most standard, neutral and free of regional, ethnic or socioeconomic features; it is spoken in many American movies, news, television productions, commercial advertisements, radio programs. Concerning Mexican Spanish pronunciation, the language spoken in university auditoriums, theatre, and mass media is also accepted as the standard accent.

In the next subsection we present the inventory of most frequent allophones of American English (AE) and Mexican Spanish in their standard accents mentioned above, indicating phonemes as well since they are a commonly used tool in L2 pronunciation teaching. Phoneme symbols are given in forward slashes and allophone symbols are put in brackets. Allophones which are pronounced exactly as phonemes are not given, so the sound pronounced as a phoneme can be viewed as the principal allophone of the phoneme. After each phoneme followed by an example word, we give only those most common allophones which acquire additional articulatory and auditory features and thus differ to various degrees from the principal allophone.

The AE and MS phonemes are grouped according to two usual categories, i.e., vowels and consonants; each category is given in its respective subsection. Within each subsection, the phonemes are ordered according to their characteristics, not according to language. This is done with the purpose to show similarities and differenced between AE and MS phonemes and allophones.

We compiled this inventory of phonemes and allophones based on our study of the state of the art works on English and Spanish phonology and phonetics by Whitley [23], Avery and Ehrlich [1], Edwards [5], Quilis [19], Moreno de Alba [15], Pineda, Castellanos, Cuétara, Galescu, Juárez, Llisterri, Pérez and Villaseñor [18].

6.1 Inventory of AE and MS Vowel Phonemes and Allophones

6.1.1. Vowels

- MS high-front /i/ as in *ipo* [ˈipo], nasalized [ĩ] as in *instante* [ĩnˈstan̪te] and *mimo* [ˈmĩmo], palatal semi-consonant [j] as in *pasión* [paˈsjon], palatal semi-vowel [i̯] as in *aire* [ˈai̯re].
- AE high-front tense unrounded /i/ as in *neat* [nit], diphthongized [iɪ] as in *flee* [fli͜ɪ], diphthongized [iə] as in *seal* [siəl], reduced [ə] or [ɪ] as in *revise* [rəˈvaɪ͡z] or [rɪˈvaɪz], lengthened [iː] as in *bee* [biː], semi-lengthened [iˑ] as in *been* [biˑn], shortened [i] as in *beat* [bit].
- AE lower high-front lax unrounded /ɪ/ as in *bit* [bit], reduced [ə] as in *chalice* [ˈtʃæləs], lengthened [ɪː] as in *carrying* [ˈkærɪːŋ].

- MS mid-front /e/ as in *este* [ˈeste], nasalized [ẽ] as in *entre* [ˈẽn̪tre], *nene* [ˈnẽne].
- AE mid-front tense unrounded /e/ as in *ate* [et], diphthongized [eɪ] as in *take* [teɪk], diphthongized and lengthened [eːɪ] as in *say* [seːɪ], diphthongized and semi-lengthened [eˑɪ] as in *name* [neˑim], diphthongized and shortened [eɪ] as in *lake* [leɪk], [i] or [ɪ] as in *Monday* [ˈmʌndɪ].
- AE lower mid-front lax unrounded /ɛ/ as in *get* [gɛt], diphthongized, r-colored and lengthened [ɛːɚ] as in *tear* [tʰ ɛːɚ], diphthongized, r-colored and semi-lengthened [ɛˑɚ] as in *scared* [ˈsk⁼ɛˑɚd], diphthongized, r-colored and shortened [ɛɚ] as in *scarce* [sk⁼ɛɚs], triphthongized [eɪə] as in *jail* [dʒeɪəl].
- AE low-front lax unrounded /æ/ as in *bat* [bæt], lengthened [æ] as in *bad* [bæːd].
- MS low-central /a/ as in *papa* [ˈpapa], nasalized [ã] as in *ambos* [ˈãmbos].
- AE lower mid-to-back central lax unrounded /ʌ/ as in *above* [əˈbʌv], [ɛ] as in *such* [sɛtʃ], [ɪ] as in *just* [dʒɪst].
- AE neutral mid-central lax unstressed unrounded /ə/ as in *above* [əˈbʌv], lower high-front lax unrounded [ɪ] as in *telephone* [ˈtelɪfon].
- AE mid-central r-colored tense /ɝ/ as in perk [pʰɝk], lengthened [ɝː] as in *sir* [sɝː], semi-lengthened [ɝˑ] as in *learn* [lɝˑn], shortened [ɝ] as in *thirst* [θɝst].
- AE mid-central r-colored lax /ɚ/ as in herder [ˈhɝdɚ], r-dropped [ə] as in *motherly* [ˈmʌðəlɪ].
- AE high-back tense rounded close /u/ as in *boot* [but], diphthongized [uə] as in *stool* [stuəl], diphthongized [uʊ] as in *do it* [ˈduʊɪt], reduced [ʊ] or [ə] as in *to own* [tʊˈon], *to go* [təˈgo], lengthened [uː] as in *blue* [bluː], semi-lengthened [uˑ] as in *food* [fuˑd], shortened [u] as in *loop* [lup].
- AE high-back lax rounded /ʊ/ as in *book* [bʊk], reduced [ʌ] or [ə] as in *would* [wʌd] or [wəd].
- MS mid-back /o/ as in *oso* [ˈoso], nasalized [õ] as in *hombre* [ˈõmbre] or *mono* [ˈmõno].
- AE mid-back tense rounded close /o/ as in *owed* [od], diphthongized [oʊ] as in *go* [goʊ], reduced [ə] as in *window* [ˈwɪndə], diphthongized and lengthened [oːʊ] as in *no* [noːʊ], diphthongized and semi-lengthened [oˑʊ] as in *load* [loˑʊd], diphthongized and shortened [oʊ] as in *coat* [kʰoʊt].
- AE low mid-back lax rounded open /ɔ/ as in *bought* [bɔt], lengthened [ɔː] as in *law* [lɔː], semi-lengthened [ɔˑ] as in *dawn* [dɔˑn], shortened [ɔ] as in *thought* [θɔt], lowered [ɒ] or [ɑ] as in *cot* [kɒt] or [kɑt].
- AE low-back lax unrounded open /ɑ/ as in *pot* [pɑt], rounded [ɒ] as in *got* [gɒt], fronted [a] as in *not* [nat], fronted and rounded [ɔ] as in *father* [ˈfɔðɚ].
- MS high-back /u/ as in *pupa* [ˈpupa], nasalized [ũ] as in *un soto* [ˈũnˈsoto] or *mundo* [ˈmũn̪do], velar semi-consonant [w] as in *cuatro* [ˈkwatro], velar semi-vowel [u̯] as in *auto* [ˈau̯to].
- AE rising low-front to high-front diphthong /aɪ/ as in *kite* [kaɪt], triphthongized [aɪə] as in *I'll* [aɪəl], reduced [ə] *I don't know* [əˈdõʔˈno], lengthened [aːɪ] as in *lie* [laːɪ], semi-lengthened [aˑɪ] as in *find* [faˑɪnd], shortened [aɪ] as in *light* [laɪt], elevated [ɜɪ] as in *ice* [ɜɪs].

- AE rising low-front to high-back diphthong /aʊ/ as in *now* [naʊ], reduced [ʌʊ] as in *house* [hʌʊs].
- AE rising mid-back to high-front diphthong /ɔɪ/ as in *voice* [vɔɪs], lengthened [ɔːɪ] as in *boy* [bɔːɪ], semi-lengthened [ɔˑɪ] as in *noise* [nɔˑɪz], shortened [ɔɪ] as in *exploit* [əksˈplɔɪt].

6.1.2. Consonants

- AE voiceless bilabial stop /p/ as in *pet* [pet], /p/ with aspirated release [pʰ] as in *poke* [pʰoʊk], /p/ with unaspirated release [p⁼] as in *spot* [sp⁼ɑt], /p/ with nasal release [p̃] as in *stop 'em* [stɑp̃m̩], unreleased [p¯] as in *top* [tɑp¯], lengthened [p:] as in *stop Pete* [ˈstɑpːit], preglottalized [ʔp] as in *conception* [kənˈsɛʔpʃn].
- MS voiceless bilabial unaspirated stop /p/ as in *poco* [ˈpoko].
- AE voiced bilabial stop /b/ as in *bet* [bet], /b/ with nasal release [b̃] as in *rob him* [rɑb̃m̩], unreleased [b¯] as in *rob* [rɑb¯], lengthened [b:] as in *rob Bob* [ˈrɑbːˈbɑbː].
- MS voiced bilabial stop /b/ as in *van* [ban], approximant (spirantized) [β̞] as in *haba* [ˈaβ̞a].
- MS voiced dental stop /d/ as in *dar* [dar], approximant (spirantized) [ð̞] as in *nada* [ˈnað̞a].
- MS voiceless dental unaspirated stop /t/ as in *tío* [ˈtɪo].
- AE voiceless alveolar stop /t/ as in *ten* [ten], /t/ with aspirated release [tʰ] as in *tape* [tʰeɪp], /t/ with unaspirated release [t⁼] as in *stop* [st⁼ɒp], /t/ with nasal release [t̃] as in *button* [bʌt̃n̩], unreleased [t¯] as in *coat* [kot¯], lengthened [t:] as in *let Tim* [ˈletːˈɪm], dentalized [t̪] as in *eighth* [eɪt̪θ], flapped [ɾ] as in *letter* [ˈleɾə], preglottalized [ʔt] as in *atlas* [ˈæʔtləs], glottal stop [ʔ] as in *button* [bʌʔn], affricated (palatalized) [tʃr̥] as in *train* [tʃr̥eɪn], affricated (palatalized) [tʃ] as in *eat yet* [ˈitʃət].
- AE voiced alveolar stop /d/ as in *den* [den], /d/ with bilateral release [d‿l] as in *cradle* [kreɪd‿l], /d/ with nasal release [d̃] as in *rod 'n reel* [rɑd̃n̩ril], unreleased [d¯] as in *dad* [dæːd¯], lengthened [d:] as in *sad Dave* [ˈsæːˈdːev], dentalized [d̪] as in *width* [wɪd̪θ], flapped [ɾ] as in *ladder* [ˈlæɾə], affricated (palatalized) [dʒr] as in *drain* [dʒreɪn], affricated (palatalized) [dʒ] as in *did you* [ˈdɪdʒə].
- AE voiceless velar stop /k/ as in *cap* [kæp], /k/ with aspirated release [kʰ] as in *keep* [kʰip], /k/ with unaspirated release [k⁼] as in *skope* [sk⁼op], /k/ with bilateral release [k‿l] as in *clock* [k‿lɑk], /k/ with nasal release [k̃] as in *beacon* [bik̃n̩], unreleased [k¯] as in *take* [teɪk¯], lengthened [k:] as in *take Kim* [teɪkːɪm], preglottalized [ʔk] as in *technical* [ˈtɛʔknɪk‿l], glottal stop [ʔ] as in *bacon* [beɪʔn̩].
- MS voiced velar unaspirated stop /k/ as in *cama* [ˈkama], palatalized [kʲ] as in *queso* [ˈkʲeso].
- AE voiced velar stop /g/ as in *gap* [gæp], /g/ with bilateral release [g‿l] as in *glee* [g‿li], /g/ with nasal release [g̃] as in *pig and goat* [ˈpɪg̃n̩ˈgot],

unreleased [g¯] as in *flag* [fl̥æg¯], lengthened [g:] as in *big grapes* [ˈbɪˈgːreɪps].

- MS voiced velar stop /g/ as in *gato* [ˈgato], approximant (spirantized) [ɣ̞] as in *el gasto* [elˈɣ̞asto].
- AE voiceless labiodental fricative /f/ as in *fan* [fæn], interdental [θ] as in *trough* [trɑθ], bilabial [ɸ] as in *comfort* [ˈkʌmɸət].
- MS voiceless bilabial fricative /f/ as in *foco* [ˈfoko].
- AE voiced labiodental fricative /v/ as in *van* [væn], devoiced [v̥] as in *have to* [ˈhæv̥tə].
- MS voiceless dental fricative /s̪/ as in *Asia* [ˈas̪ja].
- AE voiceless interdental fricative /θ/ as in *thigh* [θaɪ], voiced [ð] as in *with many* [wɪðˈmenɪ].
- AE voiced interdental fricative /ð/ as in *thy* [ðaɪ], devoiced [ð̥] as in *This is not theirs* [ð̥ɪsɪz ˈnɒʔˈð̥ɛˑəz].
- AE voiceless alveolar fricative /s/ as in *sip* [sɪp], palatalized [ʃ] as in *kiss you* [ˈkɪʃju].
- MS voiceless dorosalveolar fricative /s/ as in *sol* [sol], palatalized [ʒ] as in *pues ya* [puˈeʒa], voiced [z] as in *mismo* [ˈmizmo].
- AE voiced alveolar fricative /z/ as in *zip* [zɪp], devoiced [z̥] as in *keys* [kiz̥], palatalized [ʒ] as in *as you* [æˈʒju], stopping [d] as in *business* [ˈbɪdnɪs].
- AE voiceless palatal fricative /ʃ/ as in *mesher* [ˈmeʃə].
- MS voiceless palatal fricative /ʃ/ as in *Xola* [ˈʃola].
- AE voiced palatal fricative /ʒ/ as in *measure* [ˈmeʒə], affricate [dʒ] as in *garage* [gəˈrɑdʒ].
- MS voiced dorsal palatal fricative /ɟ/ as in *yo* [ɟo].
- MS voiceless velar fricative /x/ as in *paja* [ˈpaxa].
- AE voiceless glottal fricative /h/ as in *hat* [hæt], voiced [ɦ] as in *ahead* [əˈɦed], palatalized [ç] as in *hue* [çju], /h/ with glottal release [ʔ] as in *hello* [ʔeˈləʊ], omitted [ø] as in *he has his* [hi hæz ɪz].
- AE voiceless alveo-palatal affricate /tʃ/ as in *chin* [tʃɪn].
- AE voiced alveo-palatal affricate /dʒ/ as in *gin* [dʒɪn].
- MS voiceless palatal affricate /t͡ʃ/ as in *hacha* [at͡ʃa].
- AE voiced labiovelar glide approximant /w/ as in *wed* [wed], aspirated [hw] as in *where* [hweə], devoiced [w̥] as in *twenty* [ˈtw̥entɪ].
- MS voiced alveolar thrill approximant /r/ as in *perro* [ˈpero], devoiced hushing sibilant [r̥ʃ] as in *ver* [ber̥ʃ], sibilant flap [ɾ] as in *pero* [ˈpeɾo].
- AE voiced alveopalatal liquid approximant /r/ as in *red* [red], devoiced [r̥] as in *treat* [tr̥it], flap [ɾ] as in *very* [ˈveɾɪ], retroflexed [ɻ] as in *right* [ɻaɪt], back [r̠] as in *grey* [gr̠eɪ].
- AE voiced palatal glide approximant /j/ as in *yet* [jet], omitted [ø] as in *duty* [ˈdutɪ], devoiced [j̥̊] as in *pure* [pʰj̥̊uə].
- AE voiced alveolar lateral liquid approximant /l/ as in *led* [led], light [l] as in *lease* [lis], dark, velarized [ɫ] as in *call* [kɔɫ], syllabic, also dark [ɫ̩] as in *bottle* [bɑʔɫ̩], devoiced [l̥] as in *play* [pl̥eɪ], dentalized [ɫ̥] as in *health* [hɛɫ̥θ].
- MS voiced alveolar lateral liquid approximant /l/ as in *loco* [ˈloko].

- AE voiced bilabial nasal /m/ as in *met* [met], syllabic [m̩] as in *something* [ˈsʌm̩θɪŋ], lengthened [m:] as in *some more* [sʌˈm:ɔr], labiodentalized [ɱ] as in *comfort* [ˈkʌɱfət].
- MS voiced bilabial nasal /m/ as in *más* [mas].
- MS voiced dental nasal /n̪/ as in *antes* [ˈan̪tes].
- AE voiced alveolar nasal /n/ as in *net* [net], syllabic [n̩] as in *button* [bʌʔn̩], lengthened [n:] as in *ten names* [ten:eɪmz], labildentalized [ɱ] as in *invite* [ɪɱˈvaɪt], dentalized [n̪] as in *on Thursday* [ən̪ˈθɝzde], velarized [ŋ̩] as in *income* [ˈɪŋkəm].
- MS voiced alveolar nasal /n/ as in *nene* [ˈnene], dentalized [n̪] as in *cuanto* [ˈkwan̪to], velarized [ŋ̩] as in *banco* [ˈbaŋko].
- MS voiced palatal nasal /ɲ/ as in *año* [aɲo].
- AE voiced velar nasal /ŋ/ as in *lung* [lʌŋ], syllabic [ŋ̩] as in *lock and key* [ˈlɒkŋ̩ˈki], alveolarized [n] as in *running* [ˈrʌnɪn], stop [ŋ^{k}] or [ŋ^{g}] as in *king* [kɪŋ^{g}].

6.2 Error Patterns

In this section we give error patterns (which can be called pronunciation variations or error rules as well) which may operate in English speech generated by a L2 English learner whose mother tongue is Mexican Spanish. We built the rules given below based on a comparative analysis of American English and Mexican Spanish phonemes and allophones. This analysis is theoretic; in future we plan to validate these rules empirically by means of a comparative phonetic analysis of words/texts read by American English native speakers and L2 English learners with L1 Mexican Spanish.

Errors can be made due to similarities and/or differences of the spelling rules of two languages. In this work we considered only phonetic aspects without taking into account orthographic stereotypes of MS learners of AE.

The rules are given in two subsections, one for the vowels and the other for the consonants. The rules are represented according to the following patterns: on the left-hand side of a rule an AE sound is given; then, on the right-hand side of a rule, after an arrow, the MS sounds are given which may substitute the AE sound in English pronunciation of an MS speaker. If there is more than one MS sound which can substitute an AE sound, then such MS sounds are separated by a vertical bar (|). If two or more MS sounds are used to substitute an AE allophone, these MS sounds are combined using a plus (+) symbol. If in the latter combination of MS sounds one or more sounds can vary, the variation are separated by a vertical bar (|).

In some cases, an AE sound on the left-hand side of a rule looks exactly the same as an MS sound on the ride-hand side of the rule; for example, "Voiced bilabial nasal /m/ → voiced bilabial nasal /m/" (rule No. 22 in Section 6.2.2). Speaking in practical terms of acceptable pronunciation, it can be said, than the /m/ sound is pronounced correctly by an MS speaker. However, the MS /m/ is not exactly the same as the AE /m/, since the overall position of the speech organs are different in AE and MS. In spite of that, for teaching purposes, the AE /m/ can be considered the same as the MS /m/.

6.2.1 Vowels

1. High-front tense unrounded /i/ → high-front /i/
    - diphthongized [iɪ] → high-front /i/
    - diphthongized [iə] → high-front /i/ + mid-front /e/
    - reduced [ə] or [ɪ] → mid-front /e/ or high-front /i/
    - lengthened [iː] → high-front /i/
    - semi-lengthened [iˑ] → high-front /i/
    - shortened [i] → high-front /i/
2. Lower high-front lax unrounded /ɪ/ → high-front /i/
    - reduced [ə] → mid-front /e/
    - lengthened [ɪː] → high-front /i/
3. Mid-front tense unrounded /e/ → mid-front /e/
    - diphthongized [eɪ] → mid-front /e/ + high-front /i/
    - diphthongized and lengthened [eːɪ] → mid-front /e/ + high-front /i/
    - diphthongized and semi-lengthened [eˑɪ] → mid-front /e/ + high-front /i/
    - diphthongized and shortened [eɪ] → mid-front /e/ + high-front /i/
    - [i] or [ɪ] → high-front /i/
4. Lower mid-front lax unrounded /ɛ/ → mid-front /e/
    - diphthongized, r-colored and lengthened [ɛːɚ] → mid-front /e/ + low-central /a/ + sibilant flap [ɾ]
    - diphthongized, r-colored and semi-lengthened [ɛˑɚ] → mid-front /e/ + low-central /a/ + sibilant flap [ɾ]
    - diphthongized, r-colored and shortened [ɛɚ] → mid-front /e/ + low-central /a/ + sibilant flap [ɾ]
    - triphthongized [eɪə] → mid-front /e/ + high-front /i/ + mid-front /e/
5. Low-front lax unrounded /æ/ → mid-front /e/ | low-central /a/
    - lengthened [æ] → mid-front /e/ | low-central /a/
6. Lower mid-to-back central lax unrounded /ʌ/ → low-central /a/
    - [ɛ] → mid-front /e/
    - [ɪ] → high-front /i/
7. Neutral mid-central lax unstressed unrounded /ə/ → mid-front /e/
    - lower high-front lax unrounded [ɪ] → high-front /i/
8. Mid-central r-colored tense /ɝ/ → mid-front /e/ + sibilant flap [ɾ]
    - lengthened [ɝː] → mid-front /e/ + sibilant flap [ɾ]
    - semi-lengthened [ɝˑ] → mid-front /e/ + sibilant flap [ɾ]
    - shortened [ɝ] → mid-front /e/ + sibilant flap [ɾ]
9. Mid-central r-colored lax /ɚ/ → mid-front /e/ + sibilant flap [ɾ]
    - r-dropped [ə] → mid-front /e/
10. High-back tense rounded close /u/ → high-back /u/ | velar semi-vowel [u̯]
    - diphthongized [uə] → high-back /u/ + mid-front /e/
    - diphthongized [uʊ] → high-back /u/ | velar semi-vowel [u̯]
    - reduced [ʊ] or [ə] → high-back /u/ | velar semi-vowel [u̯] or mid-front /e/

    - lengthened [uː] → high-back /u/
    - semi-lengthened [uˑ] → high-back /u/
    - shortened [u] → high-back /u/ | velar semi-vowel [u̯]
11. High-back lax rounded /ʊ/ → high-back /u/ | velar semi-vowel [u̯]
    - reduced [ʌ] or [ə] → low-central /a/ or mid-front /e/
12. Mid-back tense rounded close /o/ → mid-back /o/
    - diphthongized [oʊ] → mid-back /o/ + velar semi-vowel [u̯] | high-back /u/
    - reduced [ə] → low-central /a/ | mid-front /e/
    - diphthongized and lengthened [oːʊ] → mid-back /o/ + velar semi-vowel [u̯] | high-back /u/
    - diphthongized and semi-lengthened [oˑʊ] → mid-back /o/ + velar semi-vowel [u̯] | high-back /u/
    - diphthongized and shortened [oʊ] → mid-back /o/ + velar semi-vowel [u̯] | high-back /u/
13. Low mid-back lax rounded open /ɔ/ → mid-back /o/
    - lengthened [ɔː] → mid-back /o/
    - semi-lengthened [ɔˑ] → mid-back /o/
    - shortened [ɔ] → mid-back /o/
    - lowered [ɒ] or [ɑ] → mid-back /o/ or low-central /a/
14. Low-back lax unrounded open /ɑ/ → mid-back /o/ | low-central /a/
    - rounded [ɒ] → mid-back /o/ | low-central /a/
    - fronted [a] → low-central /a/
    - fronted and rounded [ɔ] → mid-back /o/
15. Rising low-front to high-front diphthong /aɪ/ → low-central /a/ + high-front /i/
    - triphthongized [aɪə] → low-central /a/ + high-front /i/ + mid-front /e/
    - reduced [ə] → mid-front /e/
    - lengthened [aːɪ] → low-central /a/ + high-front /i/
    - semi-lengthened [aˑɪ] → low-central /a/ + high-front /i/
    - shortened [aɪ] → low-central /a/ + high-front /i/
    - elevated [ɜɪ] → mid-front /e/ + high-front /i/
16. Rising low-front to high-back diphthong /aʊ/ → low-central /a/ + velar semi-vowel [u̯] | high-back /u/
    - reduced [ʌʊ] → low-central /a/ + velar semi-vowel [u̯] | high-back /u/
17. Rising mid-back to high-front diphthong /ɔɪ/ → mid-back /o/ + high-front /i/
    - lengthened [ɔːɪ] → mid-back /o/ + high-front /i/
    - semi-lengthened [ɔˑɪ] → mid-back /o/ + high-front /i/
    - shortened [ɔɪ] → mid-back /o/ + high-front /i/

6.2.2 Consonants

1. Voiceless bilabial stop /p/ → voiceless bilabial unaspirated stop /p/

  - /p/ with aspirated release [pʰ] → voiceless bilabial unaspirated stop /p/
  - /p/ with unaspirated release [p⁼] → voiceless bilabial unaspirated stop /p/
  - /p/ with nasal release [p̃] → voiceless bilabial unaspirated stop /p/ + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [p̃] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/)
  - unreleased [p¯] → voiceless bilabial unaspirated stop /p/ | omitted [ø]
  - lengthened [p:] → voiceless bilabial unaspirated stop /p/
  - preglottalized [ʔp] → voiceless bilabial unaspirated stop /p/

2. Voiced bilabial stop /b/ → voiced bilabial stop /b/ | approximant (spirantized) [β̞]
   - /b/ with nasal release [b̃] → voiced bilabial stop /b/ | approximant (spirantized) [β̞] + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [b̃] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/)
   - unreleased [b¯] → voiced bilabial stop /b/ | omitted [ø]
   - lengthened [b:] → voiced bilabial stop /b/ | approximant (spirantized) [β̞]
3. Voiceless alveolar stop /t/ → voiceless dental unaspirated stop /t/
   - /t/ with aspirated release [tʰ] → voiceless dental unaspirated stop /t/
   - /t/ with unaspirated release [t⁼] → voiceless dental unaspirated stop /t/
   - /t/ with nasal release [t̃] → voiceless dental unaspirated stop /t/ + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [t̃] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/)
   - unreleased [t¯] → voiceless dental unaspirated stop /t/ | omitted [ø]
   - lengthened [t:] → voiceless dental unaspirated stop /t/
   - dentalized [t̪] → voiceless dental unaspirated stop /t/
   - flapped [ɾ] → voiceless dental unaspirated stop /t/ | voiced dental stop /d/ | approximant (spirantized) [ð̞] | sibilant flap [ɾ]
   - preglottalized [ʔt] → voiceless dental unaspirated stop /t/
   - glottal stop [ʔ] → voiceless dental unaspirated stop /t/ + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [ʔ] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/)
   - affricated (palatalized) [tʃr̥] → voiceless dental unaspirated stop /t/ + devoiced hushing sibilant [r̥ʃ]
   - affricated (palatalized) [tʃ] → voiceless palatal affricate /t͡ʃ/ | voiceless dental unaspirated stop /t/ | voiceless dental unaspirated

stop /t/ + voiced dorsal palatal fricative /j/ | palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/)

4. Voiced alveolar stop /d/ → voiced dental stop /d/ | approximant (spirantized) [ð̞]
    - /d/ with bilateral release [d‿l] → voiced dental stop /d/ | approximant (spirantized) [ð̞] + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [d] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/) + voiced alveolar lateral liquid approximant /l/
    - /d/ with nasal release [d̃] → voiced dental stop /d/ | approximant (spirantized) [ð̞] + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [d̃] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/)
    - unreleased [d⁻] → voiced dental stop /d/ | approximant (spirantized) [ð̞] | omitted [ø]
    - lengthened [d:] → voiced dental stop /d/ | approximant (spirantized) [ð̞]
    - dentalized [d̪] → voiced dental stop /d/ | approximant (spirantized) [ð̞]
    - flapped [ɾ] → voiced dental stop /d/ | approximant (spirantized) [ð̞] | sibilant flap [ɾ]
    - affricated (palatalized) [dʒr] → voiced dental stop /d/ + palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/) + sibilant flap [ɾ]
    - affricated (palatalized) [dʒ] → voiced dental stop /d/ + palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/)
5. Voiceless velar stop /k/ → voiced velar unaspirated stop /k/ | palatalized [kʲ]
    - /k/ with aspirated release [kʰ] → voiced velar unaspirated stop /k/ | palatalized [kʲ]
    - /k/ with unaspirated release [k⁼] → voiced velar unaspirated stop /k/ | palatalized [kʲ]
    - /k/ with bilateral release [k‿l] → voiced velar unaspirated stop /k/ + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [k] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/) + voiced alveolar lateral liquid approximant /l/
    - /k/ with nasal release [k̃] → voiced velar unaspirated stop /k/ + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [k] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/)
    - unreleased [k⁻] → voiced velar unaspirated stop /k/ | omitted [ø]
    - lengthened [k:] → voiced velar unaspirated stop /k/ | palatalized [kʲ]

   - preglottalized [ʔk] → voiced velar unaspirated stop /k/ | palatalized [kʲ]
   - glottal stop [ʔ] → voiced velar unaspirated stop /k/ | omitted [ø]

6. Voiced velar stop /g/ → voiced velar stop /g/ | approximant (spirantized) [ɣ̞]
   - /g/ with bilateral release [g‿l] → voiced velar stop /g/ | approximant (spirantized) [ɣ̞] + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [g] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/) + voiced alveolar lateral liquid approximant /l/
   - /g/ with nasal release [g̃] → voiced velar stop /g/ | approximant (spirantized) [ɣ̞] + (optional: a reduced vowel similar to the MS vowel used to read the respective vowel letter following [g̃] in a given word or word combination, if any, otherwise a reduced vowel similar to the MS vowels /e/ or /a/)
   - unreleased [g¯]→ voiced velar stop /g/ | approximant (spirantized) [ɣ̞] | omitted [ø]
   - lengthened [g:] → voiced velar stop /g/ | approximant (spirantized) [ɣ̞]
7. Voiceless labiodental fricative /f/ → voiceless bilabial fricative /f/
   - interdental [θ] → voiceless bilabial fricative /f/ | voiceless dental unaspirated stop /t/
   - bilabial [ɸ] → voiceless bilabial fricative /f/
8. Voiced labiodental fricative /v/ → voiced bilabial stop /b/ | approximant (spirantized) [β̞]
   - devoiced [y̥] → voiceless bilabial fricative /f/
9. Voiceless interdental fricative /θ/ → voiceless dental unaspirated stop /t/ | voiceless bilabial fricative /f/
   - voiced [ð] → approximant (spirantized) [ð̞] | voiced dental stop /d/
10. Voiced interdental fricative /ð/ → approximant (spirantized) [ð̞] | voiced dental stop /d/
    - devoiced [ð̥] → voiceless dental unaspirated stop /t/ | voiceless bilabial fricative /f/
11. Voiceless alveolar fricative /s/ → voiceless dorosalveolar fricative /s/ | voiceless dental fricative /s̪/
    - palatalized [ʃ] → voiceless palatal fricative /ʃ/
12. Voiced alveolar fricative /z/ → voiceless dorosalveolar fricative /s/ | voiceless dental fricative /s̪/ | palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/) | voiced [z] (allophone of voiceless dorosalveolar fricative /s/)
    - devoiced [z̥] → voiceless dorosalveolar fricative /s/ | voiceless dental fricative /s̪/
    - palatalized [ʒ] → palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/)
    - stopping [d] → voiced dental stop /d/ | approximant (spirantized) [ð̞]

13. Voiceless palatal fricative /ʃ/ → voiceless palatal fricative /ʃ/
14. Voiced palatal fricative /ʒ/ → palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/)
    - affricate [dʒ] → voiced dental stop /d/ + palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/)
15. Voiceless glottal fricative /h/ → voiceless velar fricative /x/
    - voiced [ɦ] → approximant (spirantized) [ɣ̞] | voiceless velar fricative /x/
    - palatalized [ç] → voiceless velar fricative /x/ + voiceless palatal fricative /ʃ/ | voiced dorsal palatal fricative /ʝ/
    - /h/ with glottal release [ʔ] → voiceless velar fricative /x/
    - omitted [ø] → omitted [ø]
16. Voiceless alveo-palatal affricate /tʃ/ → voiceless dental unaspirated stop /t/ + voiceless palatal fricative /ʃ/ (i.e., a combination of /t/ and /ʃ/) or voiceless palatal affricate /t͡ʃ/ (i.e., a single sound /t͡ʃ/)
17. Voiced alveo-palatal affricate /dʒ/ → voiced dental stop /d/ | approximant (spirantized) [ð̞] + palatalized [ʒ] (allophone of voiceless dorosalveolar fricative /s/)
18. Voiced labiovelar glide approximant /w/ → velar semi-consonant [w] (allophone of high-back /u/)
    - aspirated [hw] → voiceless velar fricative /x/ | omitted [ø] + velar semi-consonant [w] (allophone of high-back /u/)
    - devoiced [w̥] → velar semi-consonant [w] (allophone of high-back /u/)
19. Voiced alveopalatal liquid approximant /r/ → voiced alveolar thrill approximant /r/ | sibilant flap [ɾ]
    - devoiced [r̥] → voiced alveolar thrill approximant /r/ | sibilant flap [ɾ] | devoiced hushing sibilant [r̥^{ʃ}]
    - flap [ɾ] → sibilant flap [ɾ]
    - retroflexed [ɻ] → voiced alveolar thrill approximant /r/ | sibilant flap [ɾ]
    - back [r̠] → voiced alveolar thrill approximant /r/ | sibilant flap [ɾ]
20. Voiced palatal glide approximant /j/ → voiced dorsal palatal fricative /ʝ/
    - omitted [ø] → omitted [ø] | voiced dorsal palatal fricative /ʝ/
    - devoiced [ʝ̊] → voiced dorsal palatal fricative /ʝ/ | voiceless palatal fricative /ʃ̊/
21. Voiced alveolar lateral liquid approximant /l/ → voiced alveolar lateral liquid approximant /l/
    - light [l] → voiced alveolar lateral liquid approximant /l/
    - dark, velarized [ɫ] → voiced alveolar lateral liquid approximant /l/
    - syllabic, also dark [l̩] → voiced alveolar lateral liquid approximant /l/
    - devoiced [l̥] → voiced alveolar lateral liquid approximant /l/
    - dentalized [ɫ̥] → voiced alveolar lateral liquid approximant /l/
22. Voiced bilabial nasal /m/ → voiced bilabial nasal /m/
    - syllabic [m̩] → voiced bilabial nasal /m/

   - lengthened [m:] → voiced bilabial nasal /m/
   - labiodentalized [ɱ] → voiced bilabial nasal /m/
23. Voiced alveolar nasal /n/ → voiced alveolar nasal /n/
    - syllabic [n̩] → voiced alveolar nasal /n/
    - lengthened [n:] → voiced alveolar nasal /n/
    - labildentalized [ɱ] → dentalized [n̪]
    - dentalized [n̪] → dentalized [n̪]
    - velarized [ŋ̩] → velarized [ŋ̩]
24. Voiced velar nasal /ŋ/ → velarized [ŋ̩] (allophone of voiced alveolar nasal /n/)
    - syllabic [ŋ̩] → velarized [ŋ̩] (allophone of voiced alveolar nasal /n/)
    - alveolarized [n] → voiced alveolar nasal /n/
    - stop [ŋ^k] or [ŋ^g] → voiced alveolar nasal /n/ | velarized [ŋ̩] + voiced velar stop /ɡ/ | approximant (spirantized) [ɣ̞] | voiceless velar fricative /x/

7 Conclusions and Future Work

In this article, we presented error patterns built on the basis of our comparative analysis of American English and Mexican Spanish phonemes and allophones. These error patterns (or error rules) can be applied in designing the error detection module of a Computer Assisted Pronunciation Training (CAPT) System for teaching American English pronunciation to Mexican Spanish speakers.

Since individual error detection or localization for any language in general is a very difficult computational task in the area of automatic speech recognition, our error rules can help to improve the precision of error identification in intelligent tutor systems for teaching American English pronunciation.

To the best of our knowledge, error patterns in American English speech generated by Mexican Spanish speakers has not been defined in previous work which was done mainly for Castilian-originated standard Spanish. Moreover, the state of the art analysis was done for phonemes, and in this work we performed our analysis for the most common allophones of the phonemes. It is a significant contribution to the field, as allophones, not phonemes, are sounds generated in real-life speech, and a good mastering of allophones is what produces a less accented speech of an L2 English learner.

Also, error patterns can be implemented in automatic less-accented speech generation as well as in automatic error correction systems.

In future, we plan to empirically verify the theoretically derived error patterns in this work on the material of an English learner corpus including speech generated by Mexican Spanish native speakers.

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. Cruttenden, A.: Gimson's pronunciation of English. The 8th edition. Routledge, New York (2014)
4. DeMori, R., Suen, C. Y.: New Systems and Architectures for Automatic Speech Recognition and Synthesis. Springer-Verlag NY Inc. (2012)
5. Edwards, H. T.: Applied Phonetics: the Sounds of American English. Singular Pub. Group, San Diego, CA (1997)
6. Eskenazi, M.: An overview of spoken language technology for education. Speech Communication, vol. 51(10), pp. 832–844 (2009)
7. Granger, S.: Learner English on Computer. Routledge, New York (2014)
8. Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University Press, UK (1999)
9. Hismanoğlu, M.: The integration of information and communication technology into current ELT coursebooks: a critical analysis. Procedia-Social and Behavioral Sciences, vol. 15, pp. 37–45 (2011)
10. Ito, A., Lim, Y. L., Suzuki, M. & Makino, S.: Pronunciation error detection method based on error rule clustering using a decision tree. In Ninth European Conference on Speech Communication and Technology (2005)
11. Khan, B. H.: A Comprehensive E-Learning Model. Journal of e-Learning and Knowledge Society, vol. 1, pp. 33–43 (2005)
12. Levy, M., Stockwell, G.: CALL Dimensions: Options and Issues in Computer-Assisted Language Learning. Lawrence Erlbaum Associates, Inc., NJ (2006)
13. Liakin, D.: Mobile-Assisted Learning in the Second Language Classroom. International Journal of Information Technology & Computer Science, vol. 8(2), pp. 58–65 (2013)
14. Menzel, W., Herron, D., Bonaventura, P., Morton, R.: Automatic detection and correction of non-native English pronunciations. Proceedings of INSTILL, pp. 49–56 (2000)
15. Moreno de Alba, J. G.: El español en América. Fondo de cultura económica, México (2001)
16. 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)
17. Pieraccini, R.: The voice in the machine: building computers that understand speech. MIT Press (2012)
18. 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)
19. Quilis, A.: El comentario fonológico y fonético de textos: teoría y práctica. 3a edición. Arco/Libros, S.L., Madrid (1997)
20. 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)
21. Truong, K., Neri, A., Cucchiarini, C. & Strik, H.: Automatic pronunciation error detection: an acoustic-phonetic approach. In InSTIL/ICALL Symposium (2004)

22. Weigelt, L. F., Sadoff, S. J. & Miller, J. D.: Plosive/fricative distinction: The voiceless case. The Journal of the Acoustical Society of America, vol. 87(6), pp. 2729–2737 (1990)
23. Whitley, M.S.: Spanish-English Contrasts: A Course in Spanish linguistics. Georgetown University Press, Washington, D.C. (1986)
24. Witt, S.: Use of speech recognition in Computer assisted Language Learning. PhD thesis. Department of Engineering, University of Cambridge, UK (1999)
25. Yoon, S. Y., Hasegawa-Johnson, M. & Sproat, R.: Landmark-based automated pronunciation error detection. In Interspeech, pp. 614–617 (2010)
26. Zhao, T., Hoshino, A., Suzuki, M., Minematsu, N. & Hirose, K.: Automatic Chinese pronunciation error detection using SVM trained with structural features. In SLT, pp. 473–478 (2012)