REVISTA SIGNOS. ESTUDIOS DE LINGÜÍSTICA ISSN 0718-0934
© 2017 PUCV, Chile DOI: 10.4067/S0718-09342017000200195 50(94) 195-216

Comparative analysis of American English and Mexican Spanish consonants for Computer Assisted Pronunciation Training*

Análisis comparativo de las consonantes del inglés americano y español mexicano para la enseñanza de la pronunciación del inglés asistida por computadora

Olga Kolesnikova
INSTITUTO POLITÉCNICO NACIONAL
MÉXICO
kolesolga@gmail.com

Recibido: 17-X-2014 / **Aceptado:** 30-VIII-2016

Abstract

The objective of this work is two-fold. Firstly, we aim at detecting similarities and differences between the consonant systems of two languages, namely, American English and Mexican Spanish. To achieve this, we perform a theoretic comparative analysis of consonants of the two languages at the level of both phonemes and allophones. Secondly, a possible practical usage of our results is considered; therefore, as an example of an application, we consider computer-assisted pronunciation training (CAPT) for teaching American English pronunciation to Mexican Spanish speakers. In particular, we took advantage of the results of our analysis to define some hypothetic error patterns which can be used as a starting point for diagnosing possible mispronunciations, their posterior verification, and adjustment taking into account the principles of phonotactics and empirical phonetic analysis of the English learners' speech. The latter will result in error rules to be applied in a CAPT system for error identification and generation of appropriate corrective feedback. An adequate choice of correcting techniques will improve English pronunciation acquisition and help learners to develop less accented speech. Also, similarities found between the two consonant systems make it possible to organize and present the pronunciation teaching material using a stress-free method of helping learners to adjust their speech organs to new sounds building on the phonetic habits of their first language.

Key Words: Comparative phonetics, American English consonants, Mexican Spanish consonants, English pronunciation teaching, error patterns.

Resumen

El objetivo de este trabajo es doble. En primer lugar se aspira detectar similitudes y diferencias entre los sistemas de consonantes de dos idiomas: el inglés americano y el español mexicano. Para lograrlo se realiza un análisis teórico comparativo de las consonantes de dos idiomas al nivel de fonemas y también de alófonos. En segundo lugar se contempla un posible uso práctico de los resultados obtenidos, entonces como un ejemplo de aplicación se considera el proceso de enseñanza-aprendizaje de la pronunciación del inglés americano asistido por computadora (en inglés CAPT) a los estudiantes cuya lengua materna es el español mexicano. En particular, se aprovechan los resultados del análisis realizado en la definición de algunos patrones de error hipotéticos. Estos patrones se pueden usar como el punto de partida para el diagnóstico de las posibles pronunciaciones incorrectas, su verificación y ajuste posteriores tomando en cuenta los principios de la fonotáctica y el análisis fonético empírico del habla de los aprendices del inglés. Esto por último permitirá la construcción de las reglas de error y su uso en un sistema CAPT para la identificación de errores y generación de una retroalimentación correctiva apropiada. La elección de técnicas de corrección adecuadas, mejorará la pronunciación y ayudará a los estudiantes a desarrollar el habla fluida y menos acentuada. Las similitudes encontradas entre los dos sistemas de consonantes, permiten organizar y presentar el material de la enseñanza de la pronunciación mediante un método libre de estrés, que favorece a los alumnos el ajuste de sus órganos del habla a los sonidos del inglés cuya articulación se construirá a partir de sus hábitos fonéticos de la lengua materna.

Palabras Clave: Fonética comparativa, consonantes del inglés americano, consonantes del español mexicano, enseñanza de pronunciación del inglés, patrones de error.

INTRODUCTION

Correct pronunciation is a very important aspect of second language (L2) acquisition, indispensable not only for speech generation but also for adequate listening comprehension because the articulatory and auditory systems are interrelated: a learner is hardly able to recognize a sound s/he has never produced since it is absent in the first language or L1 (Levis, 2005). However, less accented speech generation and perfect listening comprehension are included in the requirements for some jobs, for instance, operators in call centers, so it is not a rare case that a learner may need more effective training in pronunciation (Hunter & Hachimi, 2012; Lockwood, 2012).

Traditional language courses teach pronunciation and auditory recognition of L2 phonemes commonly using four basic steps:(1) presentation/explanation, (2) imitation, (3) adjustment, and (4) recognition (Celce-Murcia, Brinton & Goodwin, 2010). First, the instructor describes what position the articulatory organs must take and how they must move in order to produce the target sound or sound combination; second, the learner listens to words with the target sound and repeats them; third, the teacher provides feedback and identifies, explains, and corrects errors with relevant exercises until production of the target sound is appropriate depending on the orientation of the course and the learner's level; fourthly and finally, the learner listens

to input and discriminates between a word with the target sound and a word without it.

At step 3 (adjustment), special attention is paid to correcting the student's errors. Making first articulatory attempts, learners almost always mispronounce the target sound, especially if the phoneme they are practicing at the moment is not present in 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 an intelligent tutor model to successfully perform the task of providing relevant feedback by identifying errors in the learners' speech, explaining the causes of such errors, and offering adequate corrective exercises. Such task is possible to accomplish taking into account many linguistic, psychological, and pedagogical aspects. We believe that the primary linguistic aspect is the knowledge of similarities and differences between L1 and L2 pronunciation systems. This knowledge will help to detect learner's mispronunciations and develop adequate correcting techniques as well as to design teaching methods that anticipate and prevent possible errors.

Therefore, we posed as the objective of our work, firstly, detection of similarities and differences between the phonetic systems of two languages, namely, American English (AE) and Mexican Spanish (MS), with respect to consonants only due to space limitations of a journal article. To achieve this, we perform a theoretic comparative analysis of the consonants of the two languages at the level of both phonemes and allophones. Since allophones vary across variants of a language, we have chosen the above mentioned variants of English and Spanish. To the best of our knowledge, such analysis was not done in previous work. Our comparison is done based on the study of literature on the issues of English and Spanish phonology and phonetics published to date. Secondly, as an example of an application, we consider Computer Assisted Pronunciation Training (CAPT) for teaching American English pronunciation to Mexican Spanish speakers, and in particular, the error detection component in the CAPT model. The results of our analysis are applied in defining some hypothetic error patterns which can be used as a starting point for diagnosing possible mispronunciations, their posterior verification, and adjustment taking into account the principles of phonotactics (Park, 2013) and empirical phonetic analysis of the English learners' speech (Strange, 2011). Also, we think that the similarities found between the two consonant systems will make it possible to organize and present the pronunciation teaching material using a stress-free method of helping learners to adjust their speech organs to new sounds building on their L1 phonetic habits. In our work, we considered two examples of how such pronunciation teaching strategy can be designed.

The rest of the paper is organized as follows. In Section 1 we review existing pronunciation training systems, consider the basic structure of their underlying

intelligent tutor model, and discuss current approaches to error detection. We argue that error patterns are a feasible method to facilitate individual error identification. Section 2specifies our methodology, Section 3 contains a detailed comparative description of AE and MS consonants at the level of phonemes and allophones. In Section 4 we propose error patterns, and in Section 5, consider their usage in the pronunciation acquisition process giving two examples of teaching AE consonants based on our comparative phonetic description given in Section 3. In the end of the article, we outline conclusions and future work.

1. Computer assisted pronunciation training and error detection

Today, Computer Assisted Language Learning (CALL) in general and Computer Assisted Pronunciation Training (CAPT) in particular are recognized as beneficial tools for both L2 teachers and students (Pokrivčáková, 2015). Accessibility in practically all everyday situations, flexibility, adaptability, and personalization make CALL an excellent instrument in any kind of learning: group and individual, formal and informal, stationary and mobile, in and outside classroom (Khan, 2005; Levy & Stockwell, 2006; Burbules, 2012; Liakin, 2013). A variety of commercial CAPT 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, 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. Another example of a CAPT system is the application designed by the University of Iowa Research Foundation located at http:// soundsofspeech.uiowa.edu/, see Figure 1.

Figure 1. Application developed by the University of Iowa ResearchFoundation.

Notwithstanding the impressive technological advance, intelligent tutor models still require further improvement (Strik, Truong, de Wet & Cucchiarini, 2009; Hismanoglu & Hismanoglu, 2011). The capacity of detecting individual errors in the speech of the learner and providing relevant feedback –activities performed at step 3 (adjustment and correction) of the teaching/learning process– remains an open research issue in CALL. The latter is due to a high complexity of this computational task related to automatic speech recognition (ASR) at a very fine-grained level (Yu & Deng, 2012). In this paper, we focus on this important challenge and address it by performing a comparative phonetic analysis of AE and MS consonant systems. We believe that the similarities and differences found between AE and MS consonant phonemes and allophones as the result of our analysis can be applied to facilitate the individual error detection process by predicting possible mispronunciations. Our results can also be used in the process of teaching AE consonants to MS speakers by developing strategies which anticipate and prevent possible errors. In what follows we discuss the basic elements of an intelligent tutor model (Section 1.1) and then review some existing individual error detection methods (Section 1.2).

1.1. The basic structure of an intelligent tutor model

The basic elements of an intelligent tutor model include tutor, leaner, domain, speech processing, and error detection (Swartz & Yazdani, 2012). These components perform activities which together comprise the L2 teaching-learning process.

The tutor simulates the activities of an 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 student's prior history;
- 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 his/her errors by means of speech processing and error detection, generating necessary feedback, and selecting appropriate corrective drills;
- evaluate the learner's performance;
- store the student's scores and error history.

The learner component models the human learner of English; it contains the student's data base which holds the following information on his/her prior history:

- training units studied;
- scores obtained;
- errors detected during the stage of articulation training and the auditory comprehension stage.

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

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

Speech processing is responsible for recognition of the learner's speech.

Error detection component processes the recognized speech of the student and identifies pronunciation errors.

1.2. Individual error detection

In comparison with overall learner's pronunciation evaluation (the interested reader can consult (Eskenazi, 2009) for a detailed explanation of this pronunciation correctness measure), individual error detection is a much more difficult issue due to a high complexity of automatic speech recognition 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. Until now, attempting to develop better methods for individual error detection, researches have suggested a number of procedures, the most representative of which are briefly reviewed in this section.

Weigelt, Sadoff, and Miller (1990) used decision trees to discriminate between voiceless fricatives and voiceless plosives 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 put into practice by Truong, Neri, Cucchiarini and Strik (2004) in order 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. (2004) also used classifiers based on Linear Discriminant Analysis (LDA) obtaining positive results. Strik et al. (2009) performed further experiments with the method in (Weigelt et al., 1990) 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 (Truong et al., 2004).

The error detection task was studied for languages other than Dutch. Zhao, Hoshino, Suzuki, Minematsu and Hirose (2012) used Support Vector Machines with structural features to identify Chinese pronunciation errors of Japanese learners. A decision tree algorithm was used in the work of Ito, Lim, Suzuki and Makino (2005) 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, Hasegawa-Johnson, and Sproat (2010) using a combination of confidence scoring at the phone level and landmark-based Support Vector Machines. Menzel, Herron, Bonaventura and Morton (2000) used the confidence scores provided by an HMM-based speech recognizer to localize English pronunciation errors of Italian and German speakers.

However, compared to human judgment, automatic erroneous sound detection is not at all satisfactory (Strik et al., 2009). We believe that error detection rate can be improved by using error patterns as guidelines for predicting errors in learner's speech.

2. Methodology

We based our comparative analysis of the consonants of American English (AE) and Mexican Spanish (MS) and identification of their similarities and differences on a detailed study of literature on the issue of English and Spanish phonology and phonetics published to date. We chose those publications which provide a fine-grained description of the respective sound systems specifying the features of phonemes and their most frequently met allophones: Whitley (1986), Avery and Ehrlich (1992), Edwards (1997), Quilis (1997), Moreno de Alba (2001), Pineda, Castellanos, Cuétara, Galescu, Juárez, Llisterri, Pérez and Villaseñor (2010).

We paid special attention to the existing literature on the issues of teaching English pronunciation to Spanish speakers. Unfortunately, such resources are scarce. The fullest courses are 'English Phonetics and Phonology for Spanish Speakers' by Mott (2005) and 'A Course in English Phonetics for Spanish Speakers' by Finch and Ortiz Lira (1982), but they teach British English to Castilian Spanish speakers. Such books like 'Teaching English Sounds to Spanish Speakers' by Schneider (1971), 'English Pronunciation for Spanish Speakers: Vowels' by Dale (1985), 'English Pronunciation for Spanish Speakers: Consonants' by Dale and Poms (1986) teach American English, but are limited to some aspects of pronunciation and do not consider Mexican Spanish peculiarities.

Having studied the description of English and Spanish consonants in the state of the art literature mentioned above, we made their theoretic comparison and organized our observations in such a way that makes it easy to see similarities and differences of two consonant systems. The results of our work are presented in the next section.

3. Comparative description of AE and MS consonants

Each sound is described using the following order. 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 an 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 word combination) in which this allophone is used; last, we explain in what contexts and under what conditions this allophone is produced. Additionally, every example word is transcribed; its narrow transcription is given in brackets. Throughout the text we used the IPA symbols (https://www.internationalphoneticassociation.org/content/ipa-chart).

3.1. Stop consonants

AE voiceless bilabial /p/ as in pet [pet]. Allophones:

- /p/ with aspirated release [pʰ] as in 'poke' [pʰoʊk], occurs in word-initial and stressed positions;
- /p/ with unaspirated release [p̄] as in 'spot' [sp̄ɑt], occurs in consonant clusters, especially after /s/;
- /p/ with nasal release [p̃] as in 'stop 'em' [stɑp̃m̩], occurs before a syllabic nasal;
- unreleased [p⁻] as in 'to'p [tɑp⁻], occurs word-finally and in some blend positions or clusters;
- lengthened [pː] as in 'stop Pete' [ˈstɑpːit], occurs when /p/ arrests and releases adjoining syllable(s);
- preglottalized [ʔp] as in 'conception' [kənˈsɛʔpʃn], occurs syllable-finally, before nasals or obstruents.

MS voiceless bilabial unaspirated /p/ as in *poco* [ˈpoko], occurs in all environments.

AE voiced bilabial /b/ as in 'bet' [bet]. Allophones:

- /b/ with nasal release [b̃] as in 'rob him' [rɑb̃m̩], occurs before a syllabic nasal;
- unreleased [b–] as in 'rob' [rɑb–], occurs word-finally and in some blend positions or clusters;

- lengthened [b:] as in 'rob Bob' [ˈrɑbːˈbɑbː], occurs when /b/ arrests and releases adjoining syllable(s);

MS voiced bilabial /b/ as in *van* [ban]. Allophones:

- [b] as in *van* [ban], occurs after a pause (phrase-initially, word-initially) or a nasal consonant.
- approximant (spirantized) [β̞] as in *haba* [ˈaβ̞a], occurs in complementary distribution with [b].

MS voiced dental /d/ as in *dar* [dar]. Allophones:

- [d] as in *dar* [dar], occurs after a pause (phrase-initially, word-initially), a nasal consonant or /l/;
- approximant (spirantized) [ð̞] as in *nada* [ˈnað̞a], occurs in complementary distribution with [d].

MS voiceless dental unaspirated /t/ as in *tío* [ˈtɪo], occurs in all environments.

AE voiceless alveolar /t/ as in 'ten' [ten]. Allophones:

- /t/ with aspirated release [tʰ] as in 'tape' [tʰeɪp], occurs in word-initial and stressed positions;
- /t/ with unaspirated release [t̄] as in 'stop' [st̄ɒp], occurs in consonant clusters, especially after /s/;
- /t/ with nasal release [t̃] as in 'button' [bʌt̃n̩], occurs before a syllabic nasal;
- unreleased [t⁻] as in 'coat' [kot⁻], occurs word-finally and in some blend positions or clusters;
- lengthened [tː] as in 'let Tim' [ˈletːˈɪm], occurs when /t/ arrests and releases adjoining syllable(s);
- dentalized [t̪] as in 'eighth '[eɪt̪θ], occurs before an interdental;
- flapped [ɾ] as in 'lette'r [ˈleɾə], occurs intervocalically when second vowel is unstressed;
- preglottalized [ʔt] as in 'atlas' [ˈæʔtləs], occurs syllable-finally, before nasals or obstruents;
- glottal stop [ʔ] as in 'button' [bʌʔn̩], occurs before [n̩] or [l̩];
- affricated (palatalized) [tʃr̝̊] as in 'train' [tʃr̝̊eɪn], occurs word-initially before /r/;
- affricated (palatalized) [tʃ] as in 'eat yet' [ˈitʃət] occurs when /t/ is followed by /j/ + unstressed vowel.

AE voiced alveolar /d/ as in 'den' [den]. Allophones:

- /d/ with bilateral release [d‿l] as in 'cradle' [kreɪd‿l], occurs before /l/;

- /d/ with nasal release [d̃] as in 'rod 'n reel' [rɑd̃n̩ril], occurs before a syllabic nasal;
- unreleased [d⁻] as in 'dad' [dæːd⁻], occurs word-finally and in some blend positions or clusters;
- lengthened [dː] as in 'sad Dave' [ˈsæːˈdːev], occurs when /d/ arrests and releases adjoining syllable(s);
- dentalized [d̪] as in 'width' [wɪd̪θ], occurs before an interdental;
- flapped [ɾ] as in 'ladder' [ˈlæɾə], occurs intervocalically when second vowel is unstressed;
- affricated (palatalized) [dʒr] as in 'drain' [dʒreɪn], occurs word-initially before /r/;
- affricated (palatalized) [dʒ] as in 'did you' [ˈdɪdʒə], occurs when /d/ is followed by /j/ + unstressed vowel.

AE voiceless velar /k/ as in cap [kæp]. Allophones:

- /k/ with aspirated release [kʰ] as in 'keep' [kʰip], occurs in word-initial and stressed positions;
- /k/ with unaspirated release [k̼] as in 'skope' [sk̼op], occurs in consonant clusters, especially after /s/;
- /k/ with bilateral release [k‿l] as in 'clock' [k‿lɑk], occurs before /l/;
- /k/ with nasal release [k̃] as in 'beacon' [bik̃n̩], occurs before a syllabic nasal;
- unreleased [k⁻] as in 'take' [teɪk⁻], occurs word-finally and in some blend positions or clusters;
- lengthened [kː] as in 'take Kim' [teɪkːɪm], occurs when /k/ arrests and releases adjoining syllable(s);
- preglottalized [ʔk] as in 'technical' [ˈtɛʔknɪk‿l], occurs syllable-finally, before nasals or obstruents;
- glottal stop [ʔ] as in 'bacon' [beɪʔn̩], occurs before [n̩] or [l̩].

MS voiced velar unaspirated /k/ as in *cama* [ˈkama]. Allophones:

- [k] as in *casa* [ˈkasa], occurs before front vowels and in consonant clusters;
- palatalized [kʲ] as in *queso* [ˈkʲeso], occurs in complementary distribution with [k].

AE voiced velar /ɡ/ as in 'gap' [ɡæp]. Allophones:

- /ɡ/ with bilateral release [ɡ‿l] as in 'glee' [ɡ‿li], occurs before /l/;
- /ɡ/ with nasal release [g̃] as in 'pig and goat '[ˈpɪg̃n̩ˈgot], occurs before a syllabic nasal;

- unreleased [ɡ̚] as in 'flag' [flæɡ̚], occurs word-finally and in some blend positions or clusters;
- lengthened [ɡː] as in 'big grapes' [ˈbɪˈɡːreɪps], occurs when /ɡ/ arrests and releases adjoining syllable(s).

MS voiced velar /ɡ/ as in *gato* [ˈɡato]. Allophones:

- [ɡ] as in *gasto* [ˈɡasto], occurs after a pause (phrase-initially, word-initially) or a nasal consonant;
- approximant (spirantized) [ɣ̞] as in *el gasto* [elˈɣ̞asto], occurs in complementary distribution with [ɡ].

3.2. Fricative consonants

AE voiceless labiodental /f/ as in 'fan' [fæn]. Allophones:

- interdental [θ] as in 'trough' [trɑθ], occurs in certain words;
- bilabial [ɸ] as in 'comfort' [ˈkʌmɸət], occurs after a labial.

MS voiceless bilabial /f/ as in *foco* [ˈfoko], occurs in all environments.

AE voiced labiodental /v/ as in '*væn*' [væn]. Allophone:

- devoiced [v̥] as in '*have to*' [ˈhævt̬ə], occurs word-finally, before or after a voiceless consonant.

MS voiceless dental /s̪/ as in *Asia* [ˈas̪ja], occurs in all environments.

AE voiceless interdental /θ/ as in 'thigh' [θaɪ]. Allophone:

- voiced [ð] as in 'with many' [wɪðˈmenɪ], occurs in coarticulation with a voiced consonant.

AE voiced interdental /ð/ as in 'th'y [ðaɪ]. Allophone:

- devoiced [ð̥] as in 'This is not theirs [ð̥ɪsɪz ˈnɒʔˈð̥ɛːəz], occurs before and after voiceless consonants and pauses.

AE voiceless alveolar /s/ as in 'sip' [sɪp]. Allophone:

- palatalized [ʃ] as in kiss you [ˈkɪʃju], occurs before [j].

MS voiceless dorosalveolar /s/ as in *sol* [sol]. Allophones:

- palatalized [ʒ] as in *pues ya* [puˈeʒa], occurs before a palatal consonant in rapid speech;

- voiced [z] as in *mismo* [ˈmizmo], occurs intervocalically or between a vowel and a voiced consonant.

AE voiced alveolar /z/ as in 'zip' [zɪp]. Allophones:

- devoiced [z̥] as in 'keys' [kiz̥], occurs word-finally, before or after voiceless consonants;
- palatalized [ʒ] as in 'as you' [æˈʒju], occurs before /j/;
- stopping [d] as in 'busines' [ˈbɪdnɪs], occurs in selected words.

AE voiceless palatal /ʃ/ as in 'mesher' [ˈmeʃə], occurs in all positions.

MS voiceless palatal /ʃ/ as in *Xola* [ˈʃola].

AE voiced palatal /ʒ/ as in 'measure' [ˈmeʒə]. Allophone:

- affricate [dʒ] as in 'garage' [gəˈrɑdʒ], occurs in some words borrowed from French.

MS voiced dorsal palatal /j/ as in *yo* [jo], occurs at the beginning of a syllable.

MS voiceless velar /x/ as in *paja* [ˈpaxa].

AE voiceless glottal /h/ as in 'hat' [hæt]. Allophones:

- voiced [ɦ] as in 'ahead '[əˈɦed], occurs intervocalically;
- palatalized [ç] as in 'hue' [çju], occurs when produced tensely;
- /h/ with glottal release [ʔ] as in 'hello' [ʔeˈləʊ], occurs word-initially in some words;
- omitted [ø] as in 'he has his' [hi hæzɪz], occurs when unstressed.

3.3. Affricate consonants

AE voiceless alveo-palatal /tʃ/ as in 'chin' [tʃɪn].

AE voiced alveo-palatal /dʒ/ as in 'gin' [dʒɪn].

MS voiceless palatal /tʃ/ as in *hacha* [atʃa].

3.4. Approximant consonants

AE voiced labiovelar glide /w/ as in wed [wed]. Allophones:

- aspirated [hw] as in 'where' [hweə], occurs in wh-words;
- devoiced [w̥] as in 'twenty' [ˈtw̥entɪ], occurs in voiceless clusters.

MS voiced alveolar thrill /r/ as in *perro* [ˈpero]. Allophones:

- devoiced hushing sibilant [r̝̊] as in *ver* [ber̝̊], occurs word-finally mostly in female speech;
- sibilant flap [ɾ] as in *pero* [ˈpeɾo], occurs between vowels.

AE voiced alveopalatal liquid /r/ as in 'red' [red]. Allophones:

- devoiced [r̝] as in 'treat' [tr̝it], occurs in voiceless clusters;
- flap [ɾ] as in 'very' [ˈveɾɪ], occurs between vowels;
- retroflexed [ɻ] as in 'right' [ɻaɪt], occurs in selected words;
- back [r̠] as in 'grey' [gr̠eɪ], occurs before or after /g/, /k/.

AE voiced palatal glide /j/ as in 'yet' [jet]. Allophones:

- omitted [ø] as in 'duty' [ˈdutɪ], occurs after a consonant other than a stop one;
- devoiced [j̊] as in 'pure' [pʰj̊uə], occurs after a voiceless stop consonant.

AE voiced alveolar lateral liquid /l/ as in 'led' [led]. Allophones:

- light [l] as in 'lease' [lis], occurs before a vowel;
- dark, velarized [ɬ] as in 'call' [kɔɬ], occurs after a vowel;
- syllabic, also dark [l̩] as in 'bottle' [bɑʔl̩], occurs in clusters;
- devoiced [l̥] as in 'play' [pl̥eɪ], occurs in voiceless clusters;
- dentalized [l̪] as in 'health' [hɛl̪θ], occurs before /θ/, /ð/.

3.5. Nasal consonants

AE voiced bilabial /m/ as in 'met' [met]. Allophones:

- syllabic [m̩] as in 'something' [ˈsʌmm̩θɪŋ], occurs in clusters;
- lengthened [m:] as in 'some more' [sʌˈm:ɔr], occurs when arrests and releases adjoining syllable(s);
- labiodentalized [ɱ] as in 'comfort' [ˈkʌɱfət], occurs before /f/ or /v/.

MS voiced bilabial /m/ as in más [mas].

MS voiced dental /n̪/ as in *antes* [ˈan̪tes].

AE voiced alveolar /n/ as in 'net' [net]. Allophones:

- syllabic [n̩] as in 'button' [bʌʔn̩], occurs in clustes;
- lengthened [n:] as in 'ten names' [ten:eɪmz], occurs when arrests and releases adjoining syllable(s);
- labildentalized [ɱ] as in 'invite' [ɪɱˈvaɪt], occurs before /f/ or /v/;
- dentalized [n̪] as in 'on Thursday' [ən̪ˈθɝzde], occurs before /θ/, /ð/;

- velarized [ŋ̟] as in 'income' [ˈɪŋ̟kəm], occurs before /k/ or /g/.

MS voiced alveolar /n/ as in *nene* [ˈnene]. Allophones:

- dentalized [n̪] as in *cuanto*[ˈkwan̪to], occurs before /t/ or /d/;
- velarized [ŋ̟] as in *banco* [ˈbaŋ̟ko], occurs before a velar consonant.

MS voiced palatal /ɲ/ as in *año* [aɲo].

AE voiced velar /ŋ/ as in 'lun'g [lʌŋ]. Allophones:

- syllabic [ŋ̩] as in 'lock and key' [ˈlɒkŋ̩ˈki], occurs in some clusters;
- alveolarized [n] as in 'running' [ˈrʌnɪn], occurs word-finally;
- stop [ŋ^{k}] or [ŋ^{g}] as in 'king [kɪŋ^{g}], occurs in final -ing.

4. Error patterns

In this section, we propose some basic hypothetical error patterns on the phoneme level. They are derived theoretically from the results of comparing AE and MS consonant sound systems given in Section 3. Certainly, such a 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 which we present in the following three subsections, respectively, (1) substitution of an AE phoneme by an MS phoneme, (2) insertion of an MS phoneme in an AE word, and (3) deletion of an AE phoneme. There are two main reasons which explain why 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; the second reason is orthographic, when the MS reading rules are applied to AE words. For example, 'haste' may be read as [eɪst] instead of [heɪst] because the letter *h* is not pronounced in all contexts in Spanish. However, knowing that the English *h* must be pronounced, an MS learner may read it as voiceless velar /x/ instead of AE voiceless glottal /h/ since /x/ is the MS consonant most similar to the AE /h/.

In Section 4.1 substitution error patterns are shown. We put the comment "due to orthography", if an error is made for this reason. If the reason is phonetic, we offer no comment. In Section 4.2 insertion errors are listed; they are caused by the influence of MS orthographic patterns and reading rules. Section 4.3 speaks about deletion errors.

4.1. Substitution

Table 1. Substitution errors.

AE consonant	Substituted by MS consonant
Stop voiceless consonants with aspirated release [p^h], [t^h], [k^h] as in 'pound', 'pitch', 'pancake', 'teeth',' touch', 'tin', 'cake', 'cast', 'coke'	Unaspirated release [p], [t], [k]
Stop voiced bilabial /b/ as in bet [bet] used in inter-vocal positions as in 'liberal', 'debate', 'forbade', 'possibility', 'diabolical'	Approximant (spirantized) [β̞] as in *haba* [ˈaβ̞a]
Stop voiced alveolar /d/ as in den [den] used in inter-vocal positions as in 'individual', 'prejudice', prudence, intruder, tedious	Approximant (spirantized) [ð̞] as in *nada* [ˈnað̞a]
Stop voiced velar /ɡ/ as in 'gap' [ɡæp] used in non-initial position as in 'regain', 'extravagant', 'plaguing',' regard', 'agony'	Approximant (spirantized) [ɣ̞] as in *el gasto* [elˈɣ̞asto]
Fricative voiceless interdental /θ/ as in 'thigh' [θaɪ]	Stop voiceless dental unaspirated /t/ as in *tío* [ˈtɪo]
Fricative voiced interdental /ð/ as in 'thy' [ðaɪ]	Stop voiced alveolar /d/ as in *den* [den]
Fricative voiceless glottal /h/ as in 'hat' [hæt]	Fricative voiceless velar /x/ as in *paja* [ˈpaxa]
Fricative voiced labiodental /v/ as in 'van' [væn]: due to orthography	Stop voiced bilabial /b/ as in *van* [ban]
Fricative voiced alveolar /z/ as in 'zip' [zɪp]	Fricative voiceless dorosalveolar /s/ as in *sol* [sol]
Approximant voiced alveopalatal liquid /r/ as in 'red' [red]	Approximant voiced alveolar thrill /r/ as in *perro* [ˈpero]
Nasal voiced velar /ŋ/ as in 'lung' [lʌŋ]	Nasal voiced alveolar /n/ as in *nene* [ˈnene]

4.2. Insertion

Consonant insertion is a rare phenomenon; insertion errors are typical for vowels. However, consonants may be inserted primarily for orthographic reasons; one example is so-called silent consonants in AE: *b* in comb, numb, debt, 'c' in muscle, scissors, 'd' in Wednesday, sandwich, handsome, 'g' in sign, gnaw, high, reign, 'k' in knock, know, knife, 'l' in salmon, calf, talk, 'm' in mnemonic, 'n' in autumn, column, solemn, 'p' in pneumonia, psychology, receipt, 's' in island, 'w' in answer, swart, two, etc. Since these letters are read in MS, English L2 learners tend to insert the corresponding consonants.

4.3. Deletion

The phenomenon of phoneme deletion is typical for consonant sounds, especially in word final positions since the latter is typical in MS. For instance, /s/ is deleted in final position in *mas* [mas] in the combination *más rapido* [ˈma ˈrapido]. Deletion may

occur in other environments; an example of this is deletion of initial /h/ in 'haste' considered previously in the same section.

5. Error detection using patterns

Error detection and correction are very important in language learning. In the computer assisted pronunciation training models described in Section 1, the learner's errors are to be detected automatically followed by generation of relevant explanations, teaching instructions, and corrective exercises. As we mentioned in Section 1.2, automatic error detection at the level of individual sounds is a complex task which can be enhanced by error patterns.

As an example, consider the word 'jungle' [ˈdʒʌŋgl]. We suggest that two types of transcription should be stored in the phonetic database: the correct transcription and the transcription including possible erroneous sounds annotated with their probabilities; see Table 2. In case the word pronounced by the learner differs significantly from the correct version based on a pre-defined threshold, the error detection model will take into account error pattern probabilities in order to identify the concrete error.

Table 2. Vowel pronunciation errors in the word 'jungle'.

Correct	Incorrect		
[ˈdʒʌŋgl]	Transcription	Probability	Reason
	[ˈhʌŋgl]	0.50	Orthographic
	[ˈjʌŋgl]	0.20	Substitution of /dʒ/ with /j/
	[ˈʲʌŋgl]	0.20	Substitution of /dʒ/ with /ʲ/
	[ˈdjʌŋgl]	0.10	Substitution of /dʒ/ with /dj/

6. Examples of Error-Preventive AE Sound Training

In this section we give two examples of teaching AE sounds to MS speakers taking into account the information presented in Sections 3 and 4. These examples show how the results of our comparative analysis can be applied in developing error preventing methods in pronunciation training. Example 1 includes an AE sound which does not exist in MS as a phoneme, while it appears as an allophone of another phoneme. Example 2 involves an AE phoneme absent in MS on the level of both phoneme and allophone. In both examples, the 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 and 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 student's level (elementary, intermediate, advanced).In both example we refer to these three stages.

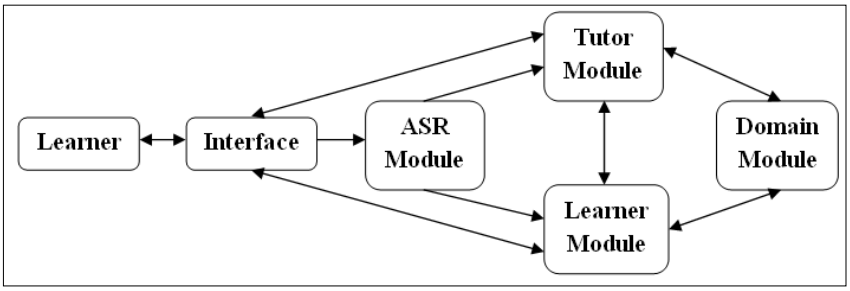

Figure 2. A model of an interactive CAPT system.

The three stages of AE phoneme training can be incorporated by a CAPT system whose main modules are shown in Figure 2. In Section 1 we mentioned the University of Iowa phonetic application (see Figure 1), in which the learner can find descriptions and visual representations of English and Spanish phonemes, however, such diagrams are located in two separate modules of that system—English and Spanish—and they have no interaction. We believe that an improved model is to be built on the contrastive interactive principle which will be more effective for training new phonemes and their allophones. We illustrate this idea by the following two examples accompanying them by the diagrams from the University of Iowa phonetic application.

Example 1

The phoneme ŋ as in lung [lʌŋ] does not exist in the MS phonemic system. Nevertheless, from Table 1 it is clear that /ŋ/ is the /n/ allophone generated in combination of /n/ with velar consonant phonemes /k/ (*banco* [ˈbaŋko]), /g/ (*pongo* [ˈpoŋgo]), /x/ (*angel* [ˈaŋxel]); therefore this allophone can be used for explaining ŋ articulation at stage 1 and initial /ŋ/ training at stage 2. The explanation may begin with the comment that /ŋ/ is a sound similar to the sound produced in MS words like *banco*, *pongo*, *angel*. These words are simple and of common usage so they are suitable for explanation, though for the training stage *angel* is not relevant because AE /ŋ/ does not combine with /h/, the phoneme most close to the MS /x/. The learner is asked to prolong the sound corresponding to the letter n in *pongo* (*pon-n-n-ngo*) thus becoming conscious of its articulation and acoustic features. Stage 1 may be accompanied by a picture (or animation) of speech organs for /ŋ/ articulation and a recording of ŋ sounding separately as well as in MS words which appear on the screen.

At stage 2, the learner is first exposed to simple AE words where the phoneme /ŋ/ appears in similar surroundings as the MS words practiced before: /ŋ/+/k/ 'drink', 'uncle', 'increase'; /ŋ/+/g/ 'singer', 'language', 'younger'. Next, /ŋ/ is introduced in combinations typical only for AE: /ŋ/+/z/ 'brings', 'thins', 'songs'; word-final /ŋ/ 'ring', 'hang', 'long', 'doing', 'nothing'. Stage 3 is devoted to auditory comprehension of AE words containing /ŋ/. Initially, the words practiced at stage 2

are presented to the learner, then other words of increasing complexity including minimal pairs (e.g. 'sin' – 'sing', 'sun' – 'sung', 'fan' – 'fang'), afterwards, short and longer phrases. At each stage, pronunciation errors are identified, explained to the learner contrasting /ŋ/ in MS and AE words, and corrected by additional exercises. Error detection process is facilitated by predicted error patterns using the results presented in Section 3. Figure 3(a) illustrates the similarity and differences of /ŋ/ and /n/.

Example 2

AE voiced alveo-palatal /dʒ/ as in 'gin' [dʒɪn] does not exist in MS as a phoneme, neither it is observed on the allophone level. However, there are MS sounds that are similar to the components of /dʒ/: dental /d/ as in *dar* [dar] and dorsal palatal /ʝ/ as in *yo* [ʝo]. So, stage 1 may begin with an explanation of this fact as well as of the differences between MS dental /d/ and AE alveolar /d/, and between MS dorsal palatal /ʝ/ and AE palatal /ʒ/ as in 'measure' [ˈmeʒə]. Then, a learner should practice both /d/ and /ʒ/ at stage 2. When the student is able to generate both AE sounds in a reasonably correct manner, s/he should be told that the two sounds must be pronounced in a connected and continuous way. The learner is to only begin articulating /d/ but instead of pronouncing it completely, the tongue must be moved down to make the /ʒ/ sound. This training stage in fact belongs to stage 1, so after practicing the components of /dʒ/, the student goes back to stage 1 to get more explanation, and then proceeds with training of /dʒ/ in various positions within words and then phrases. Figure 3(b) illustrates the similarity and differences of the respective AE and MS sounds.

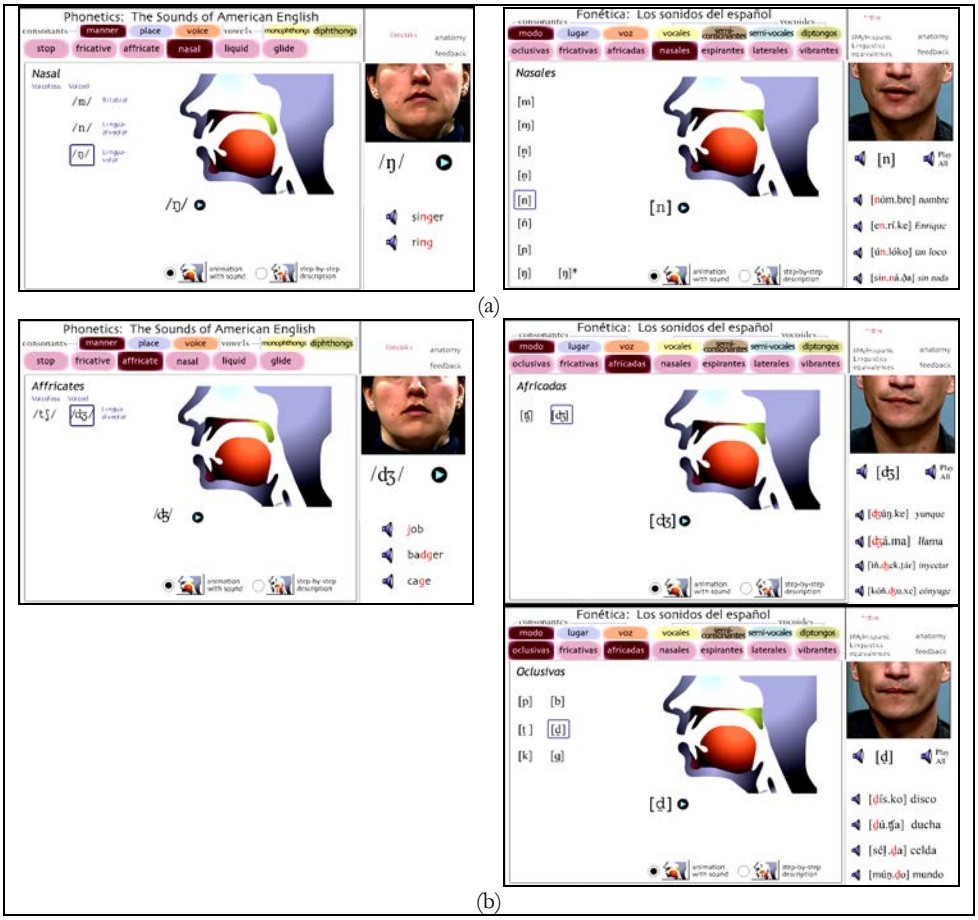

Figure 3. Similarities and differences (a) between AE /ŋ/ and MS /n/, (b) between AE /ʤ/ and MS /ʤ/ and /d/, displayed in the phonetics application of the University of Iowa Research Foundation. The AE and MS phonemes are located in two separate modules of this application.

CONCLUSIONS

In this paper, we presented the results of our detailed comparative analysis of American English (AE) and Mexican Spanish (MS) consonants on the level of both phonemes and allophones. It is a significant contribution to this research filed as such analysis had not been done in previous work. The results of our analysis are detailed contrastive descriptions of all AE and MS consonant phonemes and their most frequently observed allophones presented in such a way that it is easy to notice and explore similarities and differences in the two consonant systems.

As a possible practical application of our results we considered Computer Assister Pronunciation Training model for teaching AE pronunciation to MS speakers. In this

model, the descriptions of consonants in this article can be used for a more effective automatic individual error detection. The latter will allow for generation of a relevant feedback and presenting it to the learner. Error identification and adequate feedback generation are open research issues since the existing applications still operate on these tasks with a low precision compared to human judgment. We showed how the differences and similarities between the consonant systems of AE and MS presented in this work can be used for designing error patters to be used for mispronunciation prediction thus improving the performance of intelligent tutor applications.

Another usage of our results is development of teaching strategies which anticipate and prevent possible AE pronunciation errors in the speech of MS students. We presented two examples of how teaching articulation and auditory comprehension can be enhanced when typical error patterns are known in advance.

In future, we plan to compare the results of our theoretic phonetic analysis with errors observed empirically in learners' speech production in order to introduce modifications in error patterns proposed by us if necessary and to define a comprehensive list of error patterns. Such a list will be a valuable resource in L2 English pronunciation training via a human instructor and/or an intelligent tutor model.

REFERENCES

Avery, P. & Ehrlich, S. (1992). *Teaching American English pronunciation*. England: Oxford University Press.

Burbules N. (2012). Ubiquitous learning and the future of teaching. *Encounters on Education*, 13, 3-14.

Celce-Murcia, M., Brinton, D. & Goodwin, J. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide*. Cambridge University Press.

Dale, P. (1985). *English pronunciation for Spanish speakers: Vowels*. NJ: Prentice Hall Regents.

Dale, P. & Poms, L. (1986). *English pronunciation for Spanish speakers: Consonants*. N.J.: Prentice Hall Regents.

Edwards, H. (1997). *Applied phonetics: The sounds of American English*. San Diego, C.A.: Singular Pub. Group.

Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication*, *51*(10), 832-844.

Finch, D. & Ortiz Lira, H. (1982). *A Course in English phonetics for Spanish speakers*. London: Heinemann Educational Books Ltd.

Hismanoglu, M. & Hismanoglu, S. (2011). Internet-based pronunciation teaching: An innovative route toward rehabilitating Turkish EFL learners' articulation problems. *European Journal of Educational Studies*, *3*(1).

Hunter, M. & Hachimi, A. (2012). Talking class, talking race: Language, class, and race in the call center industry in South Africa. *Social & Cultural Geography*, *13*(6), 551-566.

Ito, A., Lim, Y., Suzuki, M. & Makino, S. (2005). Pronunciation error detection method based on error rule clustering using a decision tree. In *Proceedings of Interspeech,* 173-176.

Khan, B. (2005). A comprehensive e-learning model. *Journal of e-Learning and Knowledge Society*, 1, 33-43.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Quarterly*, *39*(3), 369-377.

Levy, M. & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning.* NJ: Lawrence Erlbaum.

Liakin, D. (2013). Mobile-assisted learning in the second language classroom. *International Journal of Information Technology* & *Computer Science*, *8*(2), 58-65.

Lockwood, J. (2012). Developing an English for specific purpose curriculum for Asian call centres: How theory can inform practice. *English for Specific Purposes*, *31*(1), 14-24.

Menzel, W., Herron, D., Bonaventura, P. & Morton, R. (2000). Automatic detection and correction of non-native English pronunciations. *Proceedings of INSTILL*, 49-56.

Moreno de Alba, J. (2001). *El español en América.* México: Fondo de Cultura Económica.

Mott, B. (2005). *English phonetics and phonology for Spanish speakers*. Barcelona: Edicions Universitat de Barcelona.

Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, *41*(2), 78-87.

Pineda, L., Castellanos, H., Cuétara, J., Galescu, L., Juárez, J., Llisterri, L., Pérez, P. & Villaseñor, L. (2010). The Corpus DIMEx100: Transcription and evaluation. *Language Resources and Evaluation*, *44*(4), 347-370.

Pokrivčáková, S. (2015). *CALL and Foreign Language Education: e-textbook for foreign language teachers*. Nitra: Constantine the Philosopher University.

Quilis, A. (1997). *El comentario fonológico y fonético de textos: Teoría y práctica*. Madrid: Arco/Libros, S.L.

Schneider, L. (1971). *Teaching English sounds to Spanish speakers*. Allied Educational Council.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of phonetics*, *39*(4), 456-466.

Strik, H., Truong, K., de Wet, F. & Cucchiarini, C. (2009). Comparing different approaches for automatic pronunciation error detection. *Speech Communication*, *51*(10), 845-852.

Swartz, M. & Yazdani, M. (Eds.). (2012). *Intelligent tutoring systems for foreign language learning: The bridge to international communication* (Vol. 80). Berlin-Heidelberg: Springer Science & Business Media.

Truong, K., Neri, A., Cucchiarini, C. & Strik, H. (2004). Automatic pronunciation error detection: An acoustic-phonetic approach. In *Proceedings of InSTIL/ICALL Symposium,* 135-138.

Weigelt, L., Sadoff, S. & Miller, J. D. (1990). Plosive/fricative distinction: The voiceless case. *The Journal of the Acoustical Society of America*, *87*(6), 2729-2737.

Whitley, M. (1986). *Spanish-English contrasts: A course in Spanish linguistics*. Washington, D.C.: Georgetown University Press.

Yoon, S., Hasegawa-Johnson, M. & Sproat, R. (2010). Landmark-based automated pronunciation error detection. *Interspeech*, 614-617.

Yu, D. & Deng, L. (2012). *Automatic speech recognition*. Berlin-Heidelberg: Springer.

Zhao, T., Hoshino, A., Suzuki, M., Minematsu, N. & Hirose, K. (2012). Automatic Chinese pronunciation error detection using SVM trained with structural features. In *Proceedings of Spoken Language Technology Workshop (SLT), IEEE,* 473-478.

* ACKNOWLEDGEMENTS

I give thanks to my God the Father and my Lord Jesus Christ for giving me life and strength to do my work. I am grateful to Instituto Politécnico Nacional, Mexico, that supported this work with grant SIP20172008 and SIP20172044, and to the Mexican Government for providing funds through SNI-CONACYT.