Pronunciation variations of Spanish-accented English spoken by young children

Hong You, Abeer Alwan

Electrical Engineering Department, UCLA
Los Angeles CA90095, USA
hyou,alwan@icsl.ucla.edu

Abe Kazemzadeh, Shrikanth Narayanan

Electrical Engineering Department, USC
Los Angeles, USA
kazemzad@usc.edu,shri@sipi.usc.edu

Abstract

When learning to speak English, non-native speakers may pronounce some English phonemes differently from native speakers. These pronunciation variations can degrade an automatic speech recognition system's performance on accented English. This paper is a first attempt to find common pronunciation variations in Spanish-accented English as spoken by young children. The analysis of pronunciation variation is performed using dynamic programming-based transcription alignment on 4500 words spoken by children 5-7 years old whose first language is Spanish. The findings are then compared with linguistic hypotheses.

1. Introduction

Automatic speech recognition (ASR) systems trained using speech from native speakers perform poorly when tested with foreign-accented speech. The mismatch between training and testing accounts for the performance degradation [8][9]. The nature of this mismatch is mainly due to pronunciation variations between native speech and foreign-accented speech. Hence, pronunciation modeling of accented speech can help improve the performance of ASR systems. A statistical model of pronunciation variations can also benefit second language learning studies.

In this paper, we summarize linguistic hypotheses of pronunciation variation of Spanish-accented English. We then compare the hypotheses with statistical analysis of 4500 words spoken by children 5-7 years old. The children's first language is Spanish.

The paper is organized as follows. Section 2 describes pronunciation variation hypotheses based on comparing Spanish and English phonetic systems. An algorithm for analyzing pronunciation variation is presented in Section 3. In Section 4, a Spanish-accented database is analyzed with the algorithm and the results are compared with the linguistic hypotheses in Section 2. Finally, Section 5 summarizes the paper.

2. Pronunciation Variation Hypotheses

Pronunciation variations of Spanish-accented English can be predicted from basic hypotheses in second language learning[6]. An important concept in second language learning is knowledge transfer, which states that the foreign speaker may habitually speak a second language using mother tongue knowledge. The knowledge transfer happens at different levels, such as phonetic transfer, grammar transfer, and orthographical transfer. In this work, we focus on pronunciation variations that are due to acoustic phonetic and orthographical transfer in Spanish-accented English.

When a native Spanish speaker listens to English, the acoustic signal is analyzed and imitated under the influence of the speaker's Spanish knowledge. If a close match to the English phoneme exists in Spanish, it is likely to be used by the Spanish speaker when producing the English phoneme. The process is what we call acoustic phonetic level transfer. There are situations when the transfer from mother tongue phonetic realization to the target language's phonetic realization is very successful. In other cases, the usage of mother tongue acoustic phonetics does not result in a close match to the canonical English phoneme, hence causing understanding difficulties. For instance, there is no sound in Spanish that is similar to the English /ih/. When learning to pronounce English /ih/, Spanish /iy/ is usually used by Spanish speakers. The differences between Spanish /iy/ and English /ih/, hence, contribute to pronunciation variation.

Orthographical knowledge transfer is another way that non-native English speakers may use when an English word is presented for them to pronounce. Usually letter-to-sound rules differ from language to language. Hence usage of the speaker's mother tongue orthographical knowledge can lead to pronunciation variations.

2.1. Consonant Pronunciation

Hypotheses of consonants' pronunciation variation in Spanish-accented English are derived by comparing the spelling rules of Spanish and English, as well as the

phonetic symbol sets of the two languages. We summarize possible pronunciation variation as predicted by [1][2][3]. Particularly, possible pronunciation variations that are within the phonemic coverage of our analysis database are listed below.

Rule 1. /v/ (vile) $\longrightarrow$ /b/ (bill) in word initial position, because spelling of the letter $\langle v \rangle$ is pronounced as /b/ in Spanish.

Rule 2. /v/ $\longrightarrow$ /f/ (fill), because /v/ doesn't exist in Spanish. Note that Spanish /f/ is acoustically similar to English /v/ except for voicing.

Rule 3. /z/ (zoo)$\longrightarrow$ /s/ (sign), because /z/ doesn't exist in Spanish and is acoustically similar to /s/. Note that one allophone of Spanish /s/ is similar to /z/.

Rule 4. /dh/ (the) $\longrightarrow$ /d/ (desk). Although /dh/ doesn't exist in Spanish, it is similar to an allophone of /d/ in Spanish. Based on sound imitation and allophone distribution, Spanish speakers tend to use /d/ as a substitute for this sound.

Rule 5. /th/ (thigh)$\longrightarrow$ /t/ (till). The voiceless counterpart of /dh/, ie. /th/, is often substituted using the voiceless counterpart of /d/, ie. /t/.

Rule 6. /r/ (right)$\longrightarrow$ /rr/ (rey in Spanish) in word initial position. Spanish /r/ has two allophones: an alveolar flap /r/, and a tongue tip trill /rr/. /rr/ occurs in Spanish at the beginning of a word.

Rule 7. /s/ $\longrightarrow$ /z/. One allophone in some Spanish dialect of /s/ is similar to English /z/.

Rule 8. /y/ (you)$\longrightarrow$ /jh/ (judge). One allophone of /y/ corresponds to English /jh/.

Rule 9. /jh/ (judge) $\longrightarrow$ /h/ (he), because spelling of the letter $\langle j \rangle$ is pronounced as /h/ in Spanish.

Rule 10. Unaspirated /p/ (pill), /t/ (till), /k/ (kill) in word initial position.

Consonants that exist in English but not in Spanish are expected to be more difficult for Spanish speakers to learn, since there is no acoustic phonetic knowledge that can provide a good transfer. Besides difficulty, larger pronunciation variations are also expected. Specifically, these consonants in English have no good Spanish counterparts: /v/, /z/, /jh/, /sh/, /h/, /dh/.

2.2. Vowel Pronunciation

The Spanish vowel system is much simpler than the English vowel system. For example, the monophthong vowels of Spanish are /i/, /e/, /a/, /o/, /u/. In addition, the duration of Spanish vowels is significantly shorter than English[4]. The following list summarizes potential pronunciation variation of Spanish-accented English vowels [1][2][3].

Rule 1. /iy/ (heed) and /ih/ (hid) are confusable, since /ih/ has no close counterpart in Spanish. Spanish /i/ is acoustically close to English /iy/, but is slightly higher in terms of tongue position.

Rule 2. /eh/ (head), /ae/ (had) are confusable, since they are close in the acoustic space. In addition, there is no close Spanish match for /ae/, while one allophone of Spanish /e/ is close to /eh/.

Rule 3. /uw/ (who) and /uh/ (hood) cover similar acoustic phonetic space as Spanish /u/, hence are confusable with each other.

Rule 4. There is no /ah/ (hud) in Spanish, hence, it tends to be pronounced as /eh/ (head).

3. Pronunciation Variation Modeling

The focus of this study is on Spanish-accented English spoken by children 5-7 years old. A broadly transcribed Spanish-accented database is utilized to statistically analyze the pronunciation variation. The TBALL database [7] contains phonemically balanced English words spoken by the children. 4500 Utterances from 9 boys and 9 girls, 5-7 years old, are transcribed. The native language of all the children is Spanish. Pronunciation variation of each word is computed using dynamic string matching, as shown in Figure 1. The transcription uses American English ARPABET symbols with additional symbols to account for Spanish-accented English's pronunciation variation [7]. The canonical pronunciations of the word list were extracted from the CMU dictionary [7]. A transcription mapping example is shown in Figure 2.

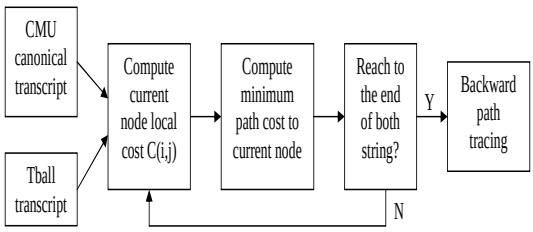

Figure 1:
Pronunciation variation analysis algorithm diagram

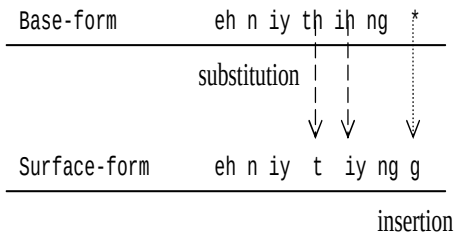

Figure 2: Dynamic string matching example

Table 1: *Variation/confusion of consonants with high likelihood analyzed from the TBALL database*

	d	f	g	jh	mb	n	s	sh	t	th	w
b	0.022	0	0	0	0.075	0	0	0	0	0	0
ch	0	0	0	0	0	0	0	0.222	0	0	0
dh	0.297	0	0.007	0	0	0	0	0	0	0.061	0
d	0	0.002	0	0.027	0	0.002	0.002	0	0.224	0.002	0
l	0.002	0	0	0	0	0.004	0	0	0	0	0.112
ng	0	0	0	0	0	0.17	0.012	0	0	0	0
t	0.039	0	0	0	0	0	0.001	0	0	0.004	0
th	0.346	0.013	0	0	0	0	0.007	0	0.092	0	0
v	0	0.216	0	0	0	0	0	0	0	0	0
y	0	0	0.054	0.08	0	0.009	0	0	0	0	0
z	0.007	0	0	0	0	0	0.736	0	0.007	0	0

4. Database Analysis

4.1. Consonant Analysis

Table 1 shows analysis results of consonants from the database. The first column is a consonant list in canonical form (expected token), while the first row is a list of consonants in possible mis-pronounced form with high likelihood (realized token). The table shows the mispronunciation likelihood from canonical phonetic symbols. We observe high substitution likelihood for consonants that don't exist in Spanish. For instance, /z/ $\longrightarrow$ /s/ 73.6%, /dh/ $\longrightarrow$ /d/ 29.7%, /v/ $\longrightarrow$ /f/ 21.6%. Alveolar stops /d/, /t/ have more pronunciation variations than other consonants. This can be attributed to the fact that the acoustic phonetic coverage of /d/ or /t/ in Spanish is larger than that in English. For example, the allophones of Spanish /d/ is as /d/ in falda, /dh/ in lado, and /d̥h/ in usted. Among the 3 allophones of Spanish /d/, /d/ is very close to English /d/, while /dh/ cover the acoustic phonetic area that is similar to /dh/ (then) in English. The voiceless allophone /d̥h/ appears mostly in word-final position. The inconsistent redistribution of /d/'s allophones can explain the large pronunciation variation observed from /th/ to /d/ (34.6%), as well as /dh/ to /d/ (29.7%). The pronunciation variation observed from /d/ to /t/ may be attributed to devoicing of consonants at word-final position.

Some of the pronunciation variation hypotheses are not observed in the TBALL data, such as /v/$\longrightarrow$/b/, /r/$\longrightarrow$/rr/, and /s/$\longrightarrow$/z/. Recall that our subjects are young children, and hence, they were exposed to English early. It is possible that these sounds get acquired earlier than others[3].

Statistical pronunciation variation analysis also leads us to some interesting new observations (denoted as O below) that have not been predicted before.

O1. /th/ $\longrightarrow$ /d/ (34.6%). This can be accounted for by the combining effects of acoustic transfer and orthographic transfer. /th/ in English is similar to an allophone of /d/ in Spanish. Based on what is heard, ie. /th/ in English, an orthographic connection of spellings of /th/ to $\langle d \rangle$ may result. Hence, this increases the likelihood of pronunciation variation from /th/ to /d/.

O2. /ch/ $\longrightarrow$ /sh/ (22.2%). Conventionally, we expect to observe pronunciation variation /sh/ $\longrightarrow$ /ch/, since /sh/ doesn't exist in Spanish. The observation from our database analysis shows that Spanish speakers have no difficulty pronouncing /sh/, and they have some tendency to omit the beginning stop /t/ of the affricate /ch/.

O3. /d/ $\longrightarrow$ /t/ (22.4%), because of the tendency to drop voicing of /d/ at the end of a word. However, this observation may not be specific to Spanish-accented English.

O4. /ng/ $\longrightarrow$ /n/ (17%) illustrate the confusion between nasal /ng/ and /n/.

In addition, some predicted pronunciation variation patterns for consonants are not observed in our analysis. For example /v/$\longrightarrow$/b/, /r/$\longrightarrow$/rr/, /s/$\longrightarrow$/z/, and the unaspirated /p/,/t/,k/. This may be explained by one or both factors listed below. First, the predicted pronunciation variation happens but at a very low likelihood. Second, our subjects are children. The influence of native language on language acquisition is less for children than for adults.

4.2. Vowel Analysis

Table 2 shows significant pronunciation variations observed in the vowels of the Tball database. Additional transcription symbols, such as /eyeh/, /ehae/, /iheh/, etc, are defined to describe non-native sounding vowels. The convention is to use two nearest vowels in the perceptual vowel space already defined in ARPABET, and the higher vowel comes first. For example, /eyeh/ is composed of /ey/ and /eh/. We observe that reduced simple vowels in English, /ah/ and /ih/,tend to have large pronunciation variation. To be specific, /ih/$\longrightarrow$/iy/ (33.4%)

Table 2: *Variation/confusion of vowels with high likelihood analyzed from the TBALL database*

	ah	eyeh	eh	ey	ehae	iy	iheh	ow	uw
ah	0	0.002	0.101	0.015	0.002	0.039	0.011	0.011	0.094
ao	0.073	0	0	0.003	0	0	0	0.321	0
aw	0.012	0	0	0	0	0	0	0.060	0
ae	0.007	0.003	0.117	0.060	0.078	0.024	0.003	0	0
eh	0.003	0.090	0	0.026	0.002	0.010	0.052	0	0
er	0.055	0	0.007	0	0	0.007	0	0	0
ih	0.006	0.003	0.055	0.006	0	0.334	0.022	0.003	0.003
uh	0.032	0	0	0	0	0	0	0	0.328

and /ah/⟶/eh/(10.1%). In addition to predicted pronunciation variations, we also observe the following possible mispronunciations from our database analysis.

O1. /ao/⟶/ow/ (32.1%). English /ao/ and /ow/ are acoustically close to each other especially for speakers with Spanish background

O2. /ae/⟶/eh/ (11.7%). There is no corresponding counterpart of English /ae/ in Spanish. Hence, a similar sound /eh/ tends to be used based on acoustic level knowledge transfer. The pronunciation variation of /ae/ is large for Spanish-accented speakers.

O3. /ah/⟶/uw/ (9.4%). Since /ah/ has no Spanish counterpart, pronunciation variations of /ah/ to front vowel /eh/ and back vowel /uw/ that appear in Spanish occur with high likelihood.

O4. /eh/⟶/eyeh/ (9%). The pronunciation of English /eh/ is similar to a sound between /ey/ and /eh/. Spanish vowel /e/ is acoustically close to English /eh/. There are two allophones of Spanish /e/. One allophone matches the English /eh/ while the other (higher vowel allophone of /e/) may be confused with English /ey/. The existence of these two allophones account for the observed pronunciation from /eh/ to /eyeh/.

5. Conclusions

In this study, we analyze pronunciation variation patterns in Spanish-accented English spoken by children. The analysis is based on comparing the acoustic phonetic systems of Spanish and English. Statistical analysis of Spanish-accented English using a database of 4500 words spoken by children 5-7 years old is carried out. The analysis shows that there are significantly large pronunciation variation for phonemes that don't exist in Spanish. Phonemes with a large number of allophones, such as /d/ in Spanish, have more pronunciation variabilities in English, which can't be accounted for by one-to-one substitution. Our analysis also shows the importance of combined data-driven and knowledge-driven pronunciation modeling. Data-driven pronunciation modeling can help to discover pronunciation variation that models the interested speaker set, while broad knowledge-driven pronunciation modeling can be used in initialization. In the future, the statistics of pronunciation variation will be employed in a speech recognizer for Spanish-accented English.

6. Acknowledgements

This work is supported by NSF grant. We thank Patti Price for providing Tball transcription guideline.

7. References

[1] Norman Coe, "Speakers of Spanish and Catalan", Learner English, Cambirdge Publishers:72-77, 1988.

[2] Avexy, P., Ehrlich S., "Teaching American English Pronunciation", Oxford Publishers, 1992.

[3] Kenworthy, J., "Teaching English Pronunciation", Longman, 1987.

[4] Stockwell, Robert P., Bowen, J. Donald, "Sounds of English and Spanish", University of Chicago Press, 1965.

[5] "Automatic speech and speaker recognition: advanced topics", edited by Chin-Hui Lee, Frank K. Soong, Kuldip K. Paliwal. Kluwer Academic Publishers, 1996.

[6] James J. Jenkins, "The learning theory approach", in "Psycholinguistics: A Survey of Theory and Research", edited by Charles E. Osgood, page 20-35, 1954.

[7] http://diana.icsl.ucla.edu/Tball/

[8] Dirk Van Compernolle, "Recognizing speech of goats, wolves, sheep and ... non-natives", Speech Communiciation, Vol 35, 2001, page 71-79.

[9] Silke Goronzy, "Robust adaptation to non-native accents in automatic speech recognition", Subseries of lecture notes in computer science, Edited by J.G. Carbonell, and J.Siekmann, Springer Publishers, 2002.