



THE SYNTHESIS OF DIALECTAL VARIATION IN ENGLISH AND SPANISH

*Mariscela Amador-Hernandez and Bathsheba J. Malsheen*

Centigram Communications Corporation  
San Jose, California, 95134 USA

ABSTRACT

At present, users of text-to-speech products can select either male or female voice, or even a multitude of different languages for various applications. However, the next generation of text-to-speech converters will need to provide additional features, such as the capability of switching from one dialect of a language to another. This paper shows that if a given language, other dialects can easily be derived from the base dialect. This paper will provide a strategy for synthesizing examples of different dialects of English and Spanish.

I. INTRODUCTION

Monolingual text-to-speech devices that offer English, German, French, Japanese, Italian, or Spanish synthetic voices, among others, have been commercially available for several years. In addition, devices which offer the capability of switching from one language to another have been developed. Some devices even provide the capability of switching from male to female voice, or from adult to child's voice. However, at present no text-to-speech device exists which allows a user to switch from one dialect to another of the same language. A specific language, for example, Spanish or English, may vary phonetically due to linguistic, geographic, or political reasons. A device which has the capability of switching dialects may not only be appropriate for specific applications, but may expedite the acceptance of synthetic speech as a means of communication.

It is well known that dialects differ from one another along many different linguistic dimensions. This presentation, however, will focus on a small number of

salient phonetic and phonological differences between two dialects of English, i.e., American and British, and two dialects of Spanish, i.e., Castilian and Latin American. It has been demonstrated that if one particular dialect is used to synthesize a given language, other dialects can be derived from the base dialect. The same kernel text-to-speech software can be modified in order to synthesize many different dialects of a language. This can be accomplished by creating and implementing a set of rules that account for variation within several different dialects of a language.

Although dialectal variation in both English and Spanish is found to affect both consonants and vowels, in this presentation only consonant variation will be demonstrated. The implementation of some of the differences between British and American English, such as the pronunciation of /t/ and /d/, final stops, and /r/ will be effected by modifying the existing

kernel text-to-speech rules for American English. Dialectal feature differences between Castilian and Latin American Spanish, such as the pronunciation of the voiceless /θ/, the palatal lateral /λ/, and the trilled /r/ will be effected by modifying the existing kernel text-to-speech rules for Castilian Spanish. Phonetic features from different dialects will be synthesized and subsequently contrasted in isolated words and short sentences.

II. DIFFERENCES BETWEEN BRITISH AND AMERICAN ENGLISH

2.1 Variants

The phonemic consonantal inventory is the same for British English (BE) as for American English (AE). There are, however, major differences in the phonological distributions of /r/ and the palatal glide /j/, and minor differences in the distributions of most other consonants. In some positions there are also important differences in the articulation of /r/, /l/

and /t/ [1].

In order to implement some of the dialectal differences found between British and American English, the existing kernel text-to-speech software rules for American English have been modified. In American English there is a rule that stipulates that /t/ becomes the flap /ɾ/ when it occurs between two vowels, the second of which is unstressed, e.g., **Betty** is pronounced as [ˈbɛɾi]. This rule does not apply to British English, where **Betty** is pronounced as [ˈbɛθi]. Therefore, it needs to be removed from a British English synthesis-by-rule system to produce native output.

Another dialectal variation which needs to be implemented is the pronunciation of the voiceless stops /p/, /t/ and /k/. These voiceless stops, when they occur in medial position before an unstressed vowel or syllabic sonorant, are in most cases pronounced with no aspiration in American English. However, they are weakly aspirated in British English, i.e., **hippie** [ˈhɪpi] vs. [ˈhɪphi], **halted** [ˈhalt@d] vs. [ˈhalt@ɪd] **biker** [ˈbajk@r] vs. [ˈbajk@hɪ].

Another dialectal variation to be implemented is the pronunciation of the voiceless alveolar stop /t/. In American English, /t/ in final position is usually not fricated, or is released as a glottal stop /ʔ/. However, in British English /t/ is always released, i.e., **eight** [ˈeɪt] or [ˈeɪʔ] vs. [ˈeɪt̚].

A final example is the American English pronunciation of the voiced, retroflex approximant /r/ that has little or no retroflexion in British English. The phonological distribution of /r/ is much more restricted in British English than in American English. British /r/ occurs only before vowels, i.e., **ferry** [ˈfɛri] vs. [ˈfəri], **carry** [ˈkɛəri] vs. [ˈkɛri]. Also British /r/ never occurs after vowels, i.e., **far** [ˈfɑr] vs. [ˈfɑ], **more** [ˈmɔr] vs. [ˈmɔ], **cure** [ˈkɛjʊr] vs. [ˈkɛjʊ].

### III. DIFFERENCES BETWEEN CASTILIAN AND LATIN AMERICAN SPANISH

#### 3.1 Variants

The consonantal phonemic inventory for Castilian Spanish is different from Latin American Spanish. Castilian Spanish has twenty consonant phonemes, but, Latin American Spanish has only eighteen [2].

To implement some of the dialectal differences found between Castilian and Latin American Spanish the existing kernel text-to-speech software rules for Castilian Spanish have been modified. In Castilian Spanish the orthographic representation of the consonant 'c' followed by the vowels 'i' and 'e', and the consonant 'z' are both

pronounced as the voiceless interdental fricative /θ/. However, in Latin American Spanish the pronunciation of these two orthographic representations has been changed. Both are pronounced as the voiceless alveolar fricative /s/. For example, **cielo** [ˈθjelo] vs. [ˈsjelo] 'sky', **cero** [ˈθero] vs. [ˈsero] 'zero', **zanja** [ˈθanja] vs. [ˈsanja] 'ditch', **zorro** [ˈθoɾo] vs. [ˈsoɾo] 'fox'.

Another dialectal difference between Castilian and Latin American Spanish is the pronunciation of the palatal lateral /ɲ/. In Latin American Spanish this sound is pronounced in many different ways. However, for our purposes we have selected

its pronunciation as a voiced palatal affricate /dʒ/ in absolute initial position or when preceded by /n/. For example, **llano** [ˈλano] vs. [ˈdʒano] 'flat', **llave** [ˈλaβe] vs. [ˈdʒaβe], **conllevar** [konλeˈβar] vs. [kondʒeˈβar] 'to bear'. In intervocalic position it is pronounced as a voiced palatal fricative /ʝ/. For example, **calle** [ˈka λ e] vs. [ˈkaʝe] 'street', **sello** [ˈse λ o] vs. [ˈseʝo] 'seal'.

A final dialectal difference to be implemented is the pronunciation of the voiced alveolar trill /r̄/. In Latin American Spanish this sound is sometimes pronounced as a voiceless alveolar trill with a strong fricative component /r̄ʃ/. For example, **rosa** [ˈr̄osa] vs. [ˈr̄osa] 'rose', **rata** [ˈr̄ata] vs. [ˈr̄ata] 'mouse', **sastre** [ˈsast̄re] vs. [ˈsast̄re] 'tailor'. In Castilian Spanish, /r̄/ is always pronounced as a trill with no frication.

### IV. COMPARISON OF DIALECTAL DIFFERENCES IN SENTENCES

#### 4.1 Sentences which demonstrate dialectal differences.

Future applications of text-to-speech may require a capability of switching from one dialect to another. In order to accomplish this, an existing kernel text-to-speech rule system that has been modified to synthesize a particular dialect of a given language may be modified to synthesize another dialect or dialects. Once the dialectal differences have been understood and implemented in the existing software, switching from one dialect to the other should be relatively straightforward. For example, a synthesis-by-rule system with the capability of switching from American English to British English should be able to output in either dialect the following sentence **Peter went around**, i.e., [ˈpɪr@r ˈwɛn aˈrãũnd] vs. [ˈpɪθ@ wɛnth aˈrãũnd]. The same capability should be possible by a system that has Castilian Spanish as its base, and Latin American Spanish as its optional dialect. This system should be able to output the

following sentence in this two dialects of Spanish, **llámenos, y le daremos mas detalles** ['λámenos j le ʎa'remos 'mas ʎ

e'ta λes] vs ['dʒámenos j le ʎa'remos 'mas ʎe'taʒes] 'give us a phone call, and we will provide you with more details'.

#### V. CONCLUSION

This paper has shown that many different dialects of a particular language can be synthesized by modifying the base dialect of a language that was initially developed. The implementation of the capability of switching from one dialect to another may be crucial to the acceptance of synthetic speech in several applications.

#### VI. REFERENCES

1. L. Bauer, J.M. Dienhart, H.H. Hartvigson, L.K. Jakobsen, 1990. American English Pronunciation. Copenhagen: Gyldendal Boghandel.
2. D. Lincoln Canfield, 1981. Spanish Pronunciation in the Americas. Chicago: University of Chicago Press.