1999-10

Model of the Classification of English Vowels by Spanish Speakers

Husaine, Fatima

Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems

http://hdl.handle.net/2144/2242

Boston University
Model of the classification of English vowels by Spanish speakers

Fatima Husain, Michiro Negishi, Michael Cohen and Frank Guenther

October, 1999


Permission to copy without fee all or part of this material is granted provided that: 1. The copies are not made or distributed for direct commercial advantage; 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and / or special permission.

Copyright © 1999

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA 02215
Model of the classification of English vowels by Spanish speakers

Fatima T. Husain¹, Michiro Negishi², Michael A. Cohen, and Frank H. Guenther

Boston University
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA, 02215

Running title: Second language vowel classification

¹. Presently at National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD.
². Presently at Psychology Department, Rutgers University, Newark, NJ.
ABSTRACT

A number of models of single language vowel classification based on formant representations have been proposed. We propose a new model that explicitly predicts vowel perception by second language (L2) learners based on the phonological map of their native language (L1). The model represents the vowels using polar coordinates in the F1-F2 formant space. Boundaries bisect the angles made by two adjacent category centroids. An L2 vowel is classified with the closest L1 vowel with a probability based on the angular difference of the L2 vowel and the L1 vowel boundary. The polar coordinate model is compared with other vowel classification models, such as the quadratic discriminant analysis method used by Hillenbrand and Gayvert [J. Speech Hear. Research, 36, 694-700, 1993] and the logistic regression analysis method adopted by Nearley [J. Phonetics, 18, 347-373, 1990]. All models were trained on Spanish vowel data and tested on English vowels. The results were compared with behavioral data obtained by Flege [Q. J. Exp. Psych., 43 A(3), 701-731 (1991)] for Spanish monolingual speakers identifying English vowels. The polar coordinate model outperformed the other models in matching its predictions most closely with the behavioral data.
1. Introduction

The phonetic categories possessed by a bilingual provide insights into formation and transformation of categories with learning. How are new phonetic and phonemic categories of a second language learned by an adult? Are the sounds of the second language (L2) represented in a different manner from the sounds of the first language (L1)? These are some of the questions explored within a modeling framework in this article. More specifically, the acquisition of phonetic categories of a second language (English) by native Spanish speakers is modeled. This special case is chosen to address bilingual research in general and to illustrate important questions regarding bilingual speech perception. In doing so, the modeling study suggests how the interference caused by the Spanish phonological map skews the perception of English vowels.

In our study we first attempted to apply the theories of second language phoneme to the Spanish and English vowel data sets. These theories are reviewed in the following section and the vowel data sets are introduced in Section 2. Since the theories do not provide a computational basis for modeling the data, two computational models (quadratic discriminant analysis and logistic regression analysis) were used to model the data. Further, as these models did not match the perceptual data adequately, a third model, the polar coordinate model, was proposed. The three models are described in Section 3. The simulations and results obtained are described in Sections 4 and 5, respectively. Issues raised by the simulation study are discussed in Section 6.

1.1. Theories of second language phoneme acquisition

Some researchers have found evidence of a common phonetic map subserving both languages spoken by a bilingual speaker (Grosjean, 1989; Blankenship, 1991; Pisoni et al., 1994; Werker, 1994; Bosch and Sebastian-Galles, 1997; Pallier et al., 1997). The shared phonetic map has been thought to cause interference when learning the phonetic categories of L2 (Best, 1994; Flege, 1991). Best’s (1994) perceptual assimilation model captures the patterns of influence that an L1 has on learning and perceiving another language. According to her model, difficulty in discriminating an L2 contrast can be predicted by examining the relationship between the L1 and L2 phonologies. The ability to discriminate two L2 sounds is determined by the discriminability of the sounds as if they were considered part of L1. For example, if each member of an L2 (phonemic) contrast is similar to a different L1 phoneme, discrimination of the L2 contrast should be very good and the two L2 phonemes assimilate into two L1 phonemes. If members of the L2 contrast are equally similar to the same L1 phoneme, discrimination is predicted to be poor. If the members of the L2 contrast assimilate to the same native phoneme but differ in the degree of similarity, then they should be relatively more discriminable. Finally, if characteristics of the L2 contrast are quite different from any native contrast, the L2 items may be as easily discriminated as are non-speech sounds and are said to be nonassimilable. Flege’s theory (Bohn and Flege, 1990, 1992; Flege, 1991, 1992) of interlingual identification extends Best’s theory to explain changes occurring in bilingual phonetic categories over time. He
Model of the classification of English vowels by Spanish speakers

views the assimilation and non-assimilation of sounds as leading to changes in the original phonetic categories. L2 sounds that are almost "identical" with L1 sounds are perceived and produced in terms of their L1 counterparts. No changes are made to the original phonological map even with experience. When the second language sounds are "similar" enough to be assimilated into the first language phonetic category, the original category expands in size to accommodate the L2 sounds. In this case, no new categories are formed immediately, but with experience, a separate category may evolve for the L2 sounds. However, if the L2 sound is "new" (i.e., if it differs substantially from any L1 phoneme) then it warrants the creation of a new category. An L2 sound approximately equidistant from two L1 categories will be identified with one or the other of these L1 categories, depending on factors like context. Flege differs from Best in what he considers to be "new" sounds. According to Best's model, new sounds are considered a separate category from the inception of exposure to L2. In Flege's viewpoint, most new categories emerge as separate entities gradually with exposure to L2.

The literature provides examples of different metrics of similarity including gestural properties (preferred by Best), perceptual distance in formant frequency space in mel or bark units with a suitable threshold (preferred by Flege), and representation of sounds by IPA (International Phonetic Alphabet) symbols. In the last case, two sounds from different languages may share the same IPA symbol but differ acoustically. Note that the above metrics are an attempt to quantify the perception of the listeners but are independent of subjective judgement. For the purposes of the current study, the terminology of Flege with regard to L2 sounds will be employed: L2 sounds can be perceived as identical, similar or new. Additionally, since we are interested in describing vowels in acoustic space, we adopt acoustic distance in mel units as a metric of similarity.

In the next section, the perception of English vowels by Spanish speakers is evaluated with respect to Flege's interlingual identification theory.

2. Perceptual data of English vowel classification by Spanish speakers

Figure 1 maps the English and Spanish vowels in the acoustic space spanned by the first two formants, F1 and F2 (data is taken from Delattre, 1969). Spanish and English vowels differ in a number of ways. A prominent difference concerns the number of vowels; English has 10 monophthong and 6 diphthong vowels (Ladefoged, 1993) while Spanish has 5 monophthong and 3 diphthong vowels (Quilis, 1981). Though Spanish vowels use the same metrics of contrast (for instance, height and front-back position of the tongue), they do not use temporal duration or "tenseness" of the vowels as a contrasting measure. Among other qualities, lax English vowels tend to be shorter in duration than their tense counterparts (Ladefoged, 1993). It should also be noted that while some English and Spanish vowels may share the same International Phonetic Association (IPA)
symbols, they are not necessarily realized in the same manner. For instance, the English /i/ and /u/ are realized with a higher tongue height than their Spanish counterparts (Quilis, 1981; Flege, 1991).

![Figure 1](image)


Table 1 lists the most frequent substitutes (Spanish vowels) used by native Spanish monolingual speakers when identifying English vowels. The two studies reported in the table are those of Scholes (1967) and Flege (1991). The listeners in the studies were asked to identify English vowels in terms of the 5 Spanish vowels and to use the “none” (or “not-a-Spanish”) label if the English vowel sounded different from any Spanish vowel. While the subjects in Scholes’s study were monolingual Spanish speakers, those in Flege’s study were both monolinguals and Spanish-English bilinguals. However, for better com-
Model of the classification of English vowels by Spanish speakers

Comparison, the table reports data only from monolingual speakers of Flege’s study. Flege also used a smaller stimulus set of only 4 English vowels compared to 8 vowels in the stimulus set used by Scholes. Scholes generated isolated, synthetic stimuli while Flege used naturally voiced stimuli in the /bVt/ syllabic context.

### TABLE 1. Vowel perception errors by Spanish learners of English

<table>
<thead>
<tr>
<th>Target English Vowel</th>
<th>Most Frequent Substitute Reported</th>
<th>Other Reported Substitutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>/i/1,2</td>
<td>/e/2, none2</td>
</tr>
<tr>
<td>/i/</td>
<td>/i/1,2</td>
<td>/e/2, none2</td>
</tr>
<tr>
<td>/ε/</td>
<td>/ε/1</td>
<td>none1</td>
</tr>
<tr>
<td>/ε/</td>
<td>/ε/1,2, /a/2</td>
<td>none1, /ε/1,2</td>
</tr>
<tr>
<td>/æ/</td>
<td>/æ/2, /ε/1</td>
<td>none2, /æ/1,2</td>
</tr>
<tr>
<td>/a/</td>
<td>/a/1</td>
<td>none1</td>
</tr>
<tr>
<td>/o/</td>
<td>/o/1</td>
<td>none1</td>
</tr>
<tr>
<td>/u/</td>
<td>/u/1</td>
<td>none1</td>
</tr>
</tbody>
</table>

1 Scholes (1967)
2 Flege (1991)

As the Spanish phonology is more dominant for the listeners in Flege’s study, they should report hearing the most similar (in terms of acoustic distance) Spanish vowel instead of the played English vowel, as hypothesized by Flege (1991). For the most part, the perceptual substitutions reported by the listeners in both studies are consistent with the acoustic data of the Spanish and English vowels (Figure 1). For instance, English /i/ is nearest to Spanish /i/ and was identified as Spanish /i/. Similarly English /a/ was heard as Spanish /a/, English /ε/ as Spanish /ε/, and so on. In the same fashion, from the acoustic data, one expects English /æ/ to be heard predominantly as Spanish /ε/ and on fewer occasions as /i/, since /i/ is closer to Spanish /ε/ than to /i/ in the acoustic map (see Figure 1). But the perceptual data from both Scholes’s and Flege’s study report just the opposite behavior. In Scholes’s study, all subjects reported that /i/ sounded like Spanish /i/. In Flege’s study the monolingual speakers identified /i/ as /i/ most of the time and only sometimes as /ε/ or as “not-a-Spanish” vowel. An English vowel somewhat isolated from the Spanish vowels, as is the case with /æ/, should be heard as a “new” or “not-a-Spanish” vowel. The listeners in the two studies differed in the extent to which they identified as a new vowel. Listeners in Scholes’s study heard it mostly as a new vowel, while listeners in Flege’s study heard it as the Spanish vowel /a/. This difference can be related to the use of natural English vowels in context in Flege’s study (as opposed to synthetic, isolated vowels in Scholes’s study) which perhaps facilitated their identification as Spanish vowels.
For the most part the theoretical predictions made by Best and Flege match the perceptual behavior of Spanish monolingual speakers. The only discrepancies involve the perception of English /ɪ/ and /æ/. Why is English /ɪ/ heard as Spanish /i/ when it is closer to Spanish /e/ in acoustic space? Why is /æ/ not perceived as a new sound by the subjects in Flege’s study? These are some of the questions that led us to conduct a modeling study of the identification of English vowels as performed by Spanish monolinguals.

3. Computational models of vowel classification

On the one hand, the two theories of bilingual phoneme perception qualitatively describe the assimilation of L2 vowels into L1 vowel categories, but fail to specify the mechanisms of representation and assimilation of the L1 and L2 vowels. On the other hand, there are computational models of vowel or phoneme identification which exactly specify the manner of monolingual vowel representation and identification but have not been used for bilingual vowel classification. Two computational models, quadratic discrimination analysis (used by Hillenbrand and Gayvert, 1993) and logistic regression analysis (used by Nearey, 1990), represent vowels based on their formant frequencies and classify vowels by using metrics other than simple acoustic distance. In our modeling study these computational models were generalized to simulate the behavior of a Spanish monolingual speaker identifying English vowels. The predictions made by these models were matched against the perceptual data reported by Flege (1991). Since these models did not adequately describe the identification of English vowels by Spanish speakers, a new model is proposed: the polar coordinate model.

The perceptual data being modeled is that of the classification of the four English vowels /ɪ, ɛ/ by Spanish monolingual speakers as reported by Flege (1991). The three models in the present study were simulated with the desire to explain the perceptual data (Flege, 1991) and, in the process, to determine an alternate metric of similarity. A basic assumption made for the modeling study was that the L1 and L2 phonologies share a common phonological map. As has been previously discussed, there is evidence for such a common map which represents both L1 and L2 phonetic categories (Pisoni et al., 1994; Werker, 1994), specifically, when the L1 is Spanish and L2 is English (Blankenship, 1991). The three models are described in this section and the data sets are described in detail in the next section.

3.1. Quadratic discriminant analysis (QDA)

The basic idea of discriminant analysis (Johnson and Wichern, 1982, pp, 504-517) is to classify based on the variance-normalized distance between the feature vector of a given token and the centroid of a particular category. Hillenbrand and Gayvert (1993) used a quadratic discriminant classification algorithm to classify vowels based on their fundamental and formant frequencies. QDA differs from the linear discriminant analysis (used by Syrdal and Gopal, 1986) mainly in the fact that in the former distances are normalized.
by individual covariance matrices for each category while in the latter, the distances are normalized by the pooled covariance matrix over all categories.

In the vowel classification study described here, a centroid and a covariance matrix were generated for each Spanish vowel category. For each English vowel token, variance-normalized distances were computed to the Spanish vowel centroids. These distances were then converted into probabilistic measures of classifying an English vowel with a Spanish vowel category. The smallest distance to a category centroid translates into the biggest probability of classifying the token with that category. See Appendix A for details of this technique.

3.2. Logistic regression analysis (LRA)

Logistic regression analysis (Christensen, 1997) is a statistical technique to analyze the relationship between stimuli (independent variables) and response categories (dependent variables); such a relationship is assumed to be logistic. Nearey (1990, 1992, 1997) used logistic regression analysis to model interactions between phonological elements in speech perception as part of his “double-weak” theory of speech perception. He examined vowel-consonant syllables where the vowels ranged from /e/ to /æ/ and the consonants from /t/ to /d/ (Nearey, 1990).

In our study, the LRA model had 15 coefficients, 3 each for the 5 Spanish vowel categories. This resulted in 14 free parameters. The coefficients represented an estimate of the Spanish phonological map. The model was then tested with English vowel production data. The probability estimates of the classification of the English vowels were calculated using the coefficients previously determined. To compensate for over-fitting the Spanish data set, the algorithm had to be modified in a non-trivial manner and in the process added one more free parameter. Details of the algorithm, including modifications made, are given in Appendix B.

3.3. Polar coordinate model

Since the two computational models described previously, QDA and LRA, failed to adequately describe the perceptual data of Spanish monolinguals classifying English vowels, we constructed a new model, the polar coordinate model.

In the polar model, the (Spanish) vowel categories are represented by their centroids in the polar coordinate system. The origin $O$ of the polar coordinate system is the mid-range of Spanish vowel in formant space and is determined as follows:

$$O = \frac{S_{F2_{max}} - S_{F2_{min}}}{2}, \frac{S_{F1_{max}} - S_{F1_{min}}}{2}$$

(1)

where the vectors described by $[S_{F2_{max}}, S_{F1_{max}}]$ and $[S_{F2_{min}}, S_{F1_{min}}]$ represent the extreme values of the Spanish vowel formant space. The category centroids are repre-
Model of the classification of English vowels by Spanish speakers

presented by \((r, \theta)\) where \(r\) represents the radial distance from the origin and \(\theta\) represents the angle made by the vowel centroid with the origin. The angle \(\theta\) represents the spectral features of the vowel. As shown in Figure 2, the angle a Spanish vowel centroid \(S_i\) makes with the x-axis through the origin \(O\), \(\theta_{S_i}\), is given by:

\[
\theta_{S_i} = \angle(S_i - O) = \tan^{-1}\left(\frac{F_{i1} - O_{F1}}{F_{i2} - O_{F2}}\right)
\]

(2)

where \(S_i\) is denoted by \([S_{iF1}, S_{iF2}]\). The radial distance \(r\) is assumed to reflect the duration (related to the laxness) of a vowel. In the Spanish vowel system, duration is not a salient cue in distinguishing vowels. Since the model is concerned with the behavior of a Spanish speaker, it assumes that the radial distance is unimportant for classification of vowels and the vowel centroid \((r, \theta)\) is represented by \(\theta\).

The model assumes that category boundaries bisect the angles made by two vowel centroids with the x-axis \((F2)\) as shown in Figure 2. The boundary \(B_{12}\) between category centroids \(S_1\) and \(S_2\) is determined as:

\[
B_{12} = \frac{\theta_{S2} - \theta_{S1}}{2} = \frac{\theta_{S1} + \theta_{S2}}{2},
\]

(3)

where \(\theta_{s1}\) and \(\theta_{s2}\) are the polar angles of the category centroids \(S_1\) and \(S_2\).

In order to classify English vowels, each English vowel token is first represented in polar coordinates. The English vowel is classified as the Spanish vowel whose centroid is
at the smallest angular distance from it (see Figure 3). Angular distances to the nearest Spanish vowel centroid and nearest boundary are then used to determine the probability of identification. The probability of identification of an English vowel \( E_j \) with a Spanish vowel \( S_i \) is calculated as

\[
p(S_i, E_j) = 0.5 + \frac{||\theta E_j - B_{i,i+1}||}{2(||S_i - B_{i,i+1}||)}
\]

and probability of identifying it with either one of the neighbors \( S_{i+1} \) or \( S_{i-1} \) is

\[
p(S_{i \pm 1}, E_j) = 1 - p(S_i|E_j)
\]

where \( E_j \) is embedded either in the sector defined by \( S_i \) and \( B_{i,i+1} \) or \( S_i \) and \( B_{i,i-1} \). In the former case, \( S_{i+1} \) is used as \( S_i \)'s neighboring Spanish vowel category and \( B_{i,i+1} \) is used as the boundary in Equations (4) and (5). In the latter case, \( S_{i-1} \) is used as the neighbor of \( S_i \) with \( B_{i,i-1} \) as the boundary in Equations (4) and (5). \( \theta E_j \) is the angle made by the English vowel with the x-axis through the origin \( O \), calculated according to the formula in Equation 2.

If the polar angle made by the English vowel exactly equals the polar angle of a Spanish vowel centroid, the probability of the English vowel being identified as that Spanish vowel equals 1 and the probability of it being identified with any other Spanish vowel equals 0 (see Figure 3). This probability falls off linearly as the angle between the English vowel and the Spanish vowel centroid increases. At the boundary between the selected Spanish vowel category and its neighbor (when the polar angle between the English vowel and boundary is 0), the probability is exactly 0.5. This results in a 50% chance of the English vowel being identified with either of the Spanish vowels lying on both sides of the boundary. Unlike the QDA and the LRA methods, which allow for multiple identifications of a vowel with varying probabilities, the polar model assumes that a vowel token can only be identified with at most two Spanish vowel categories.

4. Simulations

Simulations of the classification of English vowels by Spanish monolinguals were performed using all three models. Simulations consisted of three steps: training, testing and matching against the perceptual data.

Training: The phonological map of a Spanish monolingual was simulated using the Mexican Spanish vowel set (Godinez, 1978), since a majority of the respondents in Flege's study were Mexicans. This constituted the training of the models. The Mexican Spanish data set consists of the 5 Spanish vowel utterances of 6 male speakers who had
The probability of identifying a given English vowel \( E_i \) with any Spanish vowel category \( S_j \) is related to the polar angles \( \theta_{E_i} \) (polar angle of English vowel token \( E_i \)), \( B_{ij} \), and \( \theta_{S_j} \). If \( \theta_{E_i} \) is identical to \( \theta_{S_j} \), the probability of identification is 1. This probability decreases linearly with increasing difference between \( \theta_{E_i} \) and \( \theta_{S_j} \) and simultaneously decreasing difference between \( \theta_{E_i} \) and \( B_{ij} \). See text for details.

Testing: Once the models were trained to represent the Spanish vowel map, they were used to classify the 4 American English vowels used in Flege’s (1991) study. The American English vowel production data set consisted of multiple tokens of the English words beat, bit, bet, and bat spoken by 10 individuals (5 males and 5 females) living in Birming-
Model of the classification of English vowels by Spanish speakers

Ham, Alabama. Three tokens were collected from each individual. Each token was analyzed in order to identify the formant frequencies. Flege tabulated the average formant values (F1 and F2) for each vowel for the respective speakers. These 40 elements (10 speakers x 4 vowels) formed the English vowel production data set for the models. These 40 averaged normalized American English vowels (in mels) are plotted in Figure 4 with their IPA symbol. See Appendix C for details about the normalization process.

FIGURE 4. Normalized formant values of Spanish (indicated by *) and English vowels. The vowels are represented by IPA symbols. Spanish vowel data is taken from Godinez (1978) and the English vowel data is taken from Flege (1991).

Matching: The classification of the 4 English vowels as performed by the three models was compared with the classification by the Spanish monolinguals from Flege's (1991)
Model of the classification of English vowels by Spanish speakers

study. In Flege's study twenty Spanish monolinguals (14 Mexican, 5 El Salvadoran, and 1 Argentinian) judged the vowels /i/, /I/, /e/, and /æ/ as one of the Spanish vowel phonemes (/i/, /e/, /a/, /o/ and /u/) or as "none", if they heard the English vowel as a new or not-a-Spanish vowel. The three models were also evaluated on the basis of their classifications of Spanish vowels by matching against the Spanish vowel perception data set.

It should be noted that the QDA model normalizes acoustic distances by using the covariance of the data sets and hence it is presented with the unnormalized Spanish and English vowel data sets.

5. Simulation results

This section describes the results of the simulations conducted using the three models. The probabilistic classifications performed by the three models and the perceptual data of the classification performed by the Spanish monolinguals (Flege, 1991) are reported in Table 2.

As is evident from Table 2, the polar model outperformed the other models in matching its results most closely with the Flege (1991) perceptual data. What distinguished the polar model from the other models was how the English vowels /I/ and /æ/ were classified. Recall that contrary to the predictions of the theoretical models, the perceptual data (Scholes, 1967; Flege, 1991) suggests that English vowel /I/ is identified predominantly with Spanish vowel /i/ and that English /æ/ is treated as a new vowel. The QDA and the LRA models classified English /I/ mostly as Spanish /e/ (93% of the time by the QDA model and 60% of the time by the LRA model), in keeping with the theoretical predictions rather than with the observed perceptual data. But the polar model identified English /I/ as /I/ in 60% of the cases and as /e/ in 30%, very closely matching the numbers from Flege's data. The vowel /æ/ was identified as /a/ and as /e/ 71% and 17%, respectively in the Flege (1991) data. The polar model results compared favorably with this data, with /æ/ being identified as /a/ 73% and as /e/ 27% of the time. The LRA model classified /æ/ 58% of the time as /a/ and 41% as /e/. However, the QDA model, though classifying the majority of /æ/ as /a/ (75%), also classified it as Spanish /a/ with a probability of 35%, a result which was inconsistent with the perceptual data. It should be noted that no provision is made in any of the three models for new categories. The vowel space is divided in terms of the 5 Spanish vowel categories and the English vowels can be identified only in terms of these 5 categories.

The polar model behaved in a manner similar to the other models in classifying /I/ and /æ/. All three models classified English /I/ predominantly as Spanish /I/ and English /æ/ as Spanish /e/. They did, however, differ in the magnitude of their estimation, with the polar model most closely approximating the perceptual data in classifying /I/ and the QDA model was most like the perceptual data in classifying /æ/.
TABLE 2. Classification data produced by the three models reported in the form of probability of identifying an English vowel with a Spanish vowel. (QDA = quadratic discriminant analysis, LRA = logistic regression analysis).

<table>
<thead>
<tr>
<th>English vowel</th>
<th>Spanish vowel</th>
<th>Perceptual Data</th>
<th>QDA</th>
<th>LRA</th>
<th>Polar</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>i</td>
<td>94*</td>
<td>60</td>
<td>92</td>
<td>100</td>
</tr>
<tr>
<td>e</td>
<td></td>
<td>5</td>
<td>40</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>i</td>
<td>68*</td>
<td>6</td>
<td>24</td>
<td>60</td>
</tr>
<tr>
<td>e</td>
<td></td>
<td>19</td>
<td>93</td>
<td>60</td>
<td>35</td>
</tr>
<tr>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>æ</td>
<td>i</td>
<td>3</td>
<td></td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>e</td>
<td></td>
<td>81*</td>
<td>63</td>
<td>67</td>
<td>59</td>
</tr>
<tr>
<td>a</td>
<td></td>
<td>13*</td>
<td>6</td>
<td>23</td>
<td>35</td>
</tr>
<tr>
<td>o</td>
<td></td>
<td>31</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>æ</td>
<td>i</td>
<td>3</td>
<td></td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>e</td>
<td></td>
<td>17*</td>
<td>41</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td></td>
<td>71*</td>
<td>75</td>
<td>58</td>
<td>73</td>
</tr>
<tr>
<td>o</td>
<td></td>
<td>25</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>none</td>
<td></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: above chance rate
Figure 5 is a graphical representation of the classification performed by the polar model. The boundary lines radiate out of the origin $O$ and tessellate the vowel space into 5 categories. The English vowel token was classified with the Spanish vowel category within whose bounds it lies. The angular distances between the English vowel and the Spanish vowel centroid and boundary was converted into a probabilistic estimate of the identification of the English vowel with the particular Spanish vowel category.

**FIGURE 5.** Graphical representation of the classification performed by the polar model. Spanish vowels are represented by IPA symbols followed by the *s. English vowels are represented by IPA symbols. The vowel formants are normalized. The boundary lines are depicted by straight lines radiating out of the origin $O$. 
5.1. Evaluating the models

The models were evaluated on the basis of the proportion of variance (POV) in the actual perceptual data for which they can account. The amount of variance predicted (square of the correlation coefficient) and the number of free parameters used by three different models are tabulated in Table 3.

TABLE 3. Evaluation of the models

<table>
<thead>
<tr>
<th>Model</th>
<th>Free Parameters</th>
<th>POV(E) (English, Flege)</th>
<th>POV(S) (Spanish)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar</td>
<td>1</td>
<td>0.91</td>
<td>0.98</td>
</tr>
<tr>
<td>QDA</td>
<td>0</td>
<td>0.39</td>
<td>1</td>
</tr>
<tr>
<td>LRA</td>
<td>14+1</td>
<td>0.74</td>
<td>0.99</td>
</tr>
</tbody>
</table>

As indicated in Table 3, the polar model had just one free parameter (the origin of the polar coordinate system), the QDA had zero free parameters and the LRA model had 15 free parameters (see Appendix B). Using the metric of $POV(E)$, the polar model accounted for 91% of the data variance of the Flege perceptual data set, while the LRA model accounted for 74% and the QDA model accounted for 28% of the variance. With respect to the Spanish vowel perception data set (represented by the metric $POV(S)$), all three models were able to predict almost all of the variance, with the QDA model outperforming the others. It should be noted that three computational models were handicapped when classifying English vowels since they did not have the additional "none" category to classify English vowels that subjects of Flege's (1991) study had. The lower numbers in the $POV(E)$ column as compared to the $POV(S)$ column can partially be explained on that account. An additional point to note is that the polar model is a population model (as are the other models), modeling averages of group data. If it were to model individual data, the specifications of categories will covary with the individual data.

6. Discussion

This section discusses some of the issues raised by the modeling study of the classification of English vowels by Spanish monolinguals. In all three models described in this article, vowels are represented by their first two formants on the logarithmic mel scale. The F1 formant is said to be inversely related to the height of the vowel while the F2-F1 dimension reflects the front-back position of the tongue (Ladefoged, 1993). While formants seem to covary with articulatory positions, they are a reduced form of the acoustic information made available to the listener. Dynamic cues, such as spectral changes during the course of the utterance and the duration of the vowel, may be important in vowel recognition (Bladon, 1982; Bladon and Lindblom, 1981). Such cues were not used in the
Model of the classification of English vowels by Spanish speakers

The polar model; incorporating them would undoubtedly improve the results. There is some evidence that auditory-perceptual systems appear to behave in a logarithmic fashion (Fant, 1973). The nonlinear scale, whether mels or bark, has been shown to be more efficient in classifying vowels (Syrdal and Gopal, 1986) as compared to the linear scale (Hz). However, this claim has been disputed by Hillenbrand and Gayvert (1993).

In addition to being described in terms of high-low and front-back position of the tongue and rounding of the lips, vowels can also be described on the tense-lax dimension. In English, to some extent, tenseness of vowels covaries with vowel length: lax vowels tend to be shorter in duration than tense vowels (Ladefoged, 1993). Among Spanish speakers vowel duration is not used as a phonetic cue to discriminate between vowels (Flege 1992). But when Spanish speakers begin learning English, it is one of the first cues they master (Flege, 1992; Flege et al., 1997). The polar model is designed to behave as a Spanish monolingual and hence does not discriminate between tense and lax vowels. In fact, in the polar model tense English vowels and their corresponding lax versions are assimilated into the same Spanish vowel category. This is the case when /i/ and /ɪ/ assimilate into Spanish /i/.

Regarding the polar model, it is hypothesized that when two vowels are assimilated into a single Spanish category, the listener differentiates between the two L2 vowels on the basis of duration, and, when producing these vowels, exaggerates the temporal differences, while minimizing the spectral differences. This distinction is seen in the production data of Flege et al. (1997), where Spanish speakers reduce spectral differences when producing English /i/ and /ɪ/, while maintaining or even exaggerating the temporal differences. It is further hypothesized that if two acoustically adjacent English vowels are assimilated into two different Spanish vowels in the polar model, the spectral differences, and not the temporal difference, will be exaggerated. This is reflected in behavioral studies (Flege, 1992; Flege et al., 1997) which describe the productions of English /i/ and /ɪ/ by Spanish learners.

As noted before none of the models described in this study have provision for incorporating a "new" category. How are new categories created in the L1-L2 phonological space? Both Best’s assimilation theory and Flege’s interlingual identification theory speculate about creation of new categories. According to Best, new categories are created for sounds very different from any sound in L1, and which are as easy to discriminate from any L1 sound as are nonspeech sounds. These L2 sounds are nonassimilable and treated as “new” from the inception of L2 learning (Best et al., 1988). In his interlingual identification theory, Flege suggests that, while learning a second language, most of the L2 sounds are identified with L1 sounds, except for a few L2 sounds which form their own category. With continued exposure to L2, some of the assimilated L2 sounds begin to sound different from their L1 counterparts and eventually separate into new categories. However, this is not true for all L2 sounds, and in addition, some L2 learners are never able to distinguish between an L2 sound and its L1 category. Thus new categories for L2 sounds, according to Flege, can form both at the inception and throughout the learning period of L2. In the
polar model, category regions are defined by angular separation on a circle, and the entire vowel space is tessellated into various categories. It is hypothesized that new categories are created in the phonological map of the polar model at the regions of greatest ambiguity, i.e. the boundaries. A certain extent of region near the boundaries can be designated as a new L2 vowel region. L2 vowels near or within the new category regions will be classified some of the time as new. With experience, the new categories may grow in size. It remains to be seen, as part of further work on the polar model, whether the hypothesis of creation of new categories is supported by perceptual data.

The polar model tessellates the vowel space into pie-shaped categories. This is unlike the more common elliptical-shaped categories, seen in Flege (1991) or in Miller (1989). The elliptical categories also underlie the determining of the covariance of a category in the QDA method: determining the variances along the two dimensions is equivalent to describing an ellipse. Better results obtained with pie-shaped categories (polar model) as opposed to elliptical categories (QDA model) suggests that a pie-shape may be more natural for vowel categories. Another reason for pie-shaped categories is due to variable-rate speech. When vowels are spoken more rapidly, they are shortened and their formants change. But this change or reduction in the vowels is in a consistent direction: towards the mid-range of the vowel space. For a study of English vowel reduction see Delattre (1969). Johnson et al. (1993) have hypothesized that people hear more extreme or hyperarticulated acoustic targets while these targets are reduced in normal speech. In the hyperarticulated version, high vowels are higher, low vowels are lower, front vowels are more front and back vowels are more back. The polar model with its pie-shaped categories and radiating boundaries, lends itself naturally to such a representation with the hyper- and hypo-articulated version of a vowel falling into the same pie-slice.

7. Concluding remarks

This article proposes a model of the perception of second language vowels. The model is simple, using only the first two formants to represent the vowels. A logarithmic scale and a polar coordinate system are used to position the vowels in the F1-F2 vowel space. The center of the polar space is the mid-range of the vowel space. With this scheme, radial distance corresponds to laxness which, in the case of English, changes with vowel duration. Boundaries are drawn at the polar angle midway between two categories, at regions of greatest ambiguity. L2 vowels are classified with the closest L1 vowels with a probability proportional to the angular difference of the L2 vowel and the L1 vowel boundary. The model parameters are determined from Spanish (L1) vowels and the model is tested with English (L2) vowels. The results obtained are compared with results of other vowel classification models: quadratic discrimination analysis and logistic regression analysis. The polar model outperforms the other models in predicting perceptual data of Spanish speakers learning English. The polar model has only 1 free parameter, yet explains 91% of the variance of the perceptual data set. In contrast, the logistic regression analysis model has 14 parameters but explains only 60% of the variance of the perceptual data. The quadratic
Model of the classification of English vowels by Spanish speakers

discrimination analysis model has zero parameters (less than the polar model) but performs poorly, explaining a mere 28% of the variance in the perceptual data set.

As part of future work, we hope to extend the polar model to incorporate new categories for L2 sounds different from any L1 sounds. We hope to test the model with different L1 and L2 vowel sets and match against the perceptual data of L2 vowel identification by L1 speakers.
Model of the classification of English vowels by Spanish speakers

Acknowledgments

The authors wish to thank Jim Flege for his invaluable help in understanding English language acquisition by Spanish speakers and for reading previous versions of this paper. The authors also wish to thank Christopher Myers and Barry Horwitz for commenting on drafts of this paper. Fatima Husain and Frank Guenther are supported in part by a grant from National Institute on Deafness and other Communication Disorders (NIDCD R29 02852). Frank Guenther is also supported in part by the Alfred P. Sloan Foundation.
Appendix A

In quadratic discriminant analysis, the normalized distance \( d \) between the \( i \)th English vowel token \( E_i \) and the \( j \)th Spanish vowel category \( S_j \) was calculated as:

\[
d(E_i, S_j) = (E_i - \mu_{S_j})Cov(S_j)^{-1}(E_i - \mu_{S_j})^T
\]

where \( \mu_{S_j} \) and \( Cov(S_j) \) were the centroid and covariance matrix, respectively, of the \( j \)th Spanish vowel category. The vowel \( E_i \) is assigned to the category \( S_j \) whose centroid was the minimum distance from \( E_i \). The normalized distances were then converted into probabilistic measures as:

\[
p(S_j | E_i) = \frac{\exp(-0.5 \cdot d(E_i, S_j))}{\sum_k \exp(-0.5 \cdot d(E_i, S_k))}
\]

where \( p(S_j | E_i) \) is the probability of classifying \( E_i \) with \( S_j \). The smallest distance to a category centroid translated into the biggest probability of classifying the token with that category.

Appendix B

In the logistic regression analysis model, an evaluation function \( f \) was constructed that determined a score for a response category \( r \) given the stimulus \([F1_s, F2_s]\) according to the equation,

\[
f(r, s) = b_r + a_{1r}F1_s + a_{2r}F2_s
\]

where \( s \) is the stimulus index, \( r \) is the response category and \( F1_s \) and \( F2_s \) represent the first two formant frequencies of the stimulus. The \( b_r \) term represents bias effects that do not depend on the stimulus and the \( a_r \) term represents the importance of the stimulus relative to the bias in determining the score. A choice function, \( p(s, r) \), which was the probability of choosing category \( r \) for stimulus \( x_s \) was defined as

\[
p(s, r) = \frac{\exp(-0.5 \cdot f(r, s))}{\sum_{r'} \exp(-0.5 \cdot f(r', s))}
\]

where \( f(r, s) \) is the evaluation function outlined in Equation 8 and the summation in the denominator is over all response categories. The functional coefficients (parameters \( a \) and \( b \)) were estimated by maximum-likelihood techniques using a cost function comprising of
Model of the classification of English vowels by Spanish speakers

log-likelihood ratios (Christensen, 1997). The cost function minimizes the difference between the expected (\( \hat{p} \)) and observed probabilities (p) as shown below:

\[
ll(p, \hat{p}) = \hat{p} \cdot \log\left(\frac{p}{\hat{p}}\right) + (1 - \hat{p}) \cdot \log\left(\frac{1 - p}{1 - \hat{p}}\right)
\]  

(10)

The expected probabilities are from the data set being matched (the Spanish vowel perception data set). The observed probabilities are the classifications generated by the LRA model.

The coefficients \( b, a_1, \) and \( a_2 \) were estimated for each vowel category of the Spanish vowel data set and the classification function for English vowels was generated using these coefficients. Since there were 5 Spanish vowel categories, there were 15 coefficients (representing the Spanish phonological map) or 14 free parameters for the LRA model. The probability estimates of the classification of the English vowels were calculated using the coefficients previously determined and as defined in Equation 9.

As there are only 30 data points in the training (Spanish vowel production) data set, the algorithm over-fit the data set and resulted in poor performance on the testing (English vowel production) data set. Hence, further modifications were made to the model which consisted of the addition of a term penalizing too strong a dependence on the categories (r). This modification added an extra parameter to the model. The modified cost function was given as

\[
C = ll(p, \hat{p}) + \gamma \sum_r \left( b_r^2 + a_1^2 + a_2^2 \right)
\]  

(11)

where the \( \gamma \) term ties together the coefficients. When the \( \gamma \) term was zero, the choice function reduced to the standard log-likelihood function and was maximized by standard maximum likelihood techniques. For the present study, the optimization parameter \( \gamma \) was varied from 0 to 1 so as to maximize the proportions of variance of the data predicted by the model and optimal fit to the perceptual data (Fleget, 1991) of the English vowel classification occurred at \( \gamma = 0.2 \). The method of penalizing for over-fitting is similar to that of ridge regression which is discussed in Ryan (1997).

Appendix C

The vowel formants were converted from the linear frequency (Hz) scale to the logarithmic (mel) scale according to the equation

\[
m = \frac{1000}{\log 2} \log \left(1 + \frac{f}{1000}\right)
\]  

(12)
where $m$ is frequency in mels and $f$ is the frequency in Hz.

The vowels in the Spanish and English vowel production sets were normalized to account for speaker-specific differences. The vowels of each speaker were normalized according to the equation

$$S'_{ij} = \frac{S_{ij} - S_{jmin}}{S_{jmax} - S_{jmin}}$$

(13)

where $S_{jmin}$ and $S_{jmax}$ are the minimum and maximum formant frequencies, respectively, spoken by the $j$th speaker and $S_{ij}'$ is the normalized version of $S_{ij}$ (the $i$th vowel spoken by the $j$th speaker). $S_{ij}$ is a vector represented by $[S_{ijF1}, S_{ijF2}, S_{ijF3}]$ and $S_{jmax}$ and $S_{jmin}$ are vectors described as $[S_{F1max}, S_{F1max}]$ and $[S_{F2min}, S_{F1min}]$. Since there was no information regarding the fundamental frequencies of the Spanish speakers in the Godinez (1978) study, normalization along the lines suggested by Miller (1989) could not be undertaken.

Since the English vowel production of Flge (1991) data does not contain the full set of English phonemes, $F1$ and $F2$ range were adjusted using Peterson and Barney (1952) data as follows. For both men and women, the lowest $F2$ is that of the vowel /ɔ/ but in the test data set the lowest $F2$ was that of /æ/. Therefore it was adjusted as

$$F2_{min} = F2_{maxT} - (F2_{maxT} - F2_{minT}) \times \frac{(O_{F2PBm} - F2_{maxPB})}{(a_{F2PBm} - F2_{maxPB})}$$

(14)

where $F2_{minT}$ is the minimum value of $F2$ used for normalization, $F2_{maxT}$ is the maximum value of $F2$ in the test data, and $F2_{minT}$ is the minimum value of $F2$ in the test data. Further, from the Peterson and Barney (1952) data, $O_{F2PB}$ is the average $F2$ of /ɔ/ and $a_{F2PBm}$ is the average $F2$ of /æ/ and $F2_{maxPB}$ is the maximum value of $F2$ across all vowels.

For men, the highest $F1$ of English vowels is that of /ɑ/ but in the English vowel production data set used, the highest $F1$ is that of /æ/. Therefore the maximum value of $F1$ for men was adjusted as

$$F1_{maxm} = F1_{minTm} + (F1_{maxTm} - F1_{minTm}) \times \frac{(a_{F1PBm} - F1_{minPBm})}{(a_{F1PBm} - F1_{minPBm})}$$

(15)

where $F1_{maxm}$ is the maximum value of $F1$ used for normalization of male data, $F1_{maxTm}$ is the maximum value of $F1$ in the test data, $F1_{minTm}$ is the minimum value of $F1$ in the test data. Further from the Peterson and Barney (1952) data, $a_{F2PBm}$ is the average $F2$ of /ɔ/ and $a_{F2PBm}$ is the average $F2$ of /æ/ and $F1_{minPBmen}$ is the minimum value of $F1$ formant frequency. All values in this equation were determined from male production data.
Reference
Model of the classification of English vowels by Spanish speakers


