

The Influence of Phonetic Features on the Perception of Accented Speech

Richard Cali

Submitted to the Department of Psychology

State University of New York at New Paltz

in partial fulfillment for the degree of Masters of Liberal Arts

Abstract

When learning a second language, speakers often produce certain phonetic features of that language differently than native speakers, which results in accented speech. Between Spanish and English, there are a number of phonetic differences which may be produced differently by Spanish speaking learners of English. I focus on two specific features of Spanish-English accented speech. The first is the lack of aspiration in pre-vocalic voiceless consonants, which results in English listeners perceiving the voiced counterparts. The second is the devoicing of the post-vocalic voiced consonants, which results in English listeners perceiving the voiceless counterparts. To test whether these two features have a strong effect on the intelligibility of accented speech, participants were asked to perform a forced-choice word recognition task in which they were presented with an auditory token, spoken by a Spanish-English bilingual, and asked to identify which of two visually presented words the spoken word was. The auditory token differed on pre- and post-vocalic voicing. I predicted that English listeners' accuracy would be worse when they listened to the pre-vocalic, voiceless tokens compared to the pre-vocalic, voiced tokens and that the opposite would occur in the post-vocalic position. As predicted, results indicated that the subjects' accuracy was better in the pre-vocalic voiced condition and the post-vocalic voicing condition. This implies that these two dimensions might have a direct effect on English listeners' perception of an accent, which suggests that at some level, individual segmental qualities can result in accent perception, independent of the presence of suprasegmental features, or even other phonetic differences

Late learners of a second language (L2) generally produce speech in that language differently than native speakers, resulting in an accent. When native speakers of that language engage in conversation with these learners it can lead to difficulties in communication (Williams & Escudero, 2014). Intelligibility of speakers can be decreased (Bent & Bradlow, 2003; Derwing & Munro, 2002) based on differences in speech production from the L2 speaker. Models of second language acquisition, such as the Speech Learning Model (Flege, 1995) or the Perceptual Assimilation Model (Best, 1995) propose that native language acquisition influences one's ability to acquire a second language. In the Speech Learning Model (Flege, 1995), individuals' perception of speech becomes attuned to the specific differences between first language (L1) phones. With these contrasts in place, language learners form phonetic categories in their L1, which mentally represent the sounds of the languages, and the differences among them. When learning a new language, these phonetic categories constrain the production of phonemes in the L2 such that the language learners will either form a new, perceptually novel category or conform to an already existing category based on the phonetic dissimilarity to the phonemes of the L1. If the L2 phoneme is perceptually related to the L1 phoneme, the production of the phoneme will be more similar to the native language, and thus could be produced with a noticeable accent. In the Perceptual Assimilation Model (Best, 1995), speech sounds become assimilated into one's native category based on phonetic similarity to native phonemes in terms of their gestural properties. In other words, based on how similarly the phoneme is articulated it will be more easily assimilated into one's native category. Non-native phones can be good or poor exemplars of native phonetic segments, or completely different from any native phones. The Perceptual Assimilation

Model considers listeners' ability to discriminate between contrastive non-native pairs the basis of assimilation. Both pairs of a contrast could sound similar to different native phonemes, and would thus become assimilated to each separately in two-category assimilation. Both could be equally similar to one native phoneme and would be assimilated to one category in single-category assimilation. Finally, each could be similar to one phoneme but one member of the contrast could be a better example of that phoneme than the other, and this would result in category-goodness assimilation. The primary difference between these two models is whether accented speech is the result of gestural differences, as in the Perceptual Assimilation Model, or differences between mental categories, such as in the Speech Learning Model. The Speech Language Model suggests that phonetic differences between the two languages accounts for the differences in acquisition. In the Perceptual Assimilation Model, however, the explanation is primarily articulatory, such that the manner in which individuals produce a second language is dependent on the articulations of their native language. However, both of these models explain accent formation as a result of the influence of one's native categories on the acquisition of L2 phonemic categories.

In this study, I will examine how native listeners of English perceive Spanish-accented English. Spanish and English differ in both segmental (e.g., phonetic) and suprasegmental properties (e.g., stress placement, prosody, and rhythm [Bertran, 1999]). For instance, Spanish is a syllable-timed language with no invariable stress placement such that the stress of each sentence is variable and changes more frequently. English, on the other hand, is generally considered stress-timed, which means that stress placement in a sentence is equally spaced and invariable, as opposed to variable (Bertran, 1999).

Spanish speakers impose their native rhythm on their second language, resulting in their English speech production being accented (Fernandez, 2005). In addition, there is a difference in the placement of stress in a sentence which indicates the topic of the sentence, which is the nuclear stress placement. In Spanish, it is commonly at the end of a sentence, while in English its position can vary from sentence to sentence.

Recently there has been more research implying the importance of suprasegmental features on accent perception. Jilka (2000) suggested that intonational deviations in accent are an essential feature of accent production and perception. For instance, American English speakers have a higher intonational slope (i.e. pattern of pitch changes in a sentence) than do German English speakers. In addition, Pellegrino (2014) found that native-Italian speakers rated Chinese-Italian bilinguals as having a lower rate of speech. While suprasegmental differences have a role in the formation of accented speech in Spanish-English bilinguals, in the current study, I will restrict myself to investigating segmental differences. Specifically, I will be examining the contribution of phonetic differences.

The relative role of phonetic features in accent formation and perception has been a point of discussion for the past two decades (Pellegrino, 2014). Flege, Bohn, & Jang (1997) suggest that an individual's ability to learn second language vowels is reduced or lost when they pass a critical period. The researchers presented English vowel sounds produced by native German, Korean, Mandarin, and Spanish speakers to native English speaking participants and asked them to identify the correct phone. Their results indicated that vowel identification performance was reduced when produced by speakers with different language backgrounds. Research has also suggested that there are differences

between consonant productions as well (Flege et al., 1998). The researchers measured the voice-onset-times of [t]s produced by native English and native Spanish speakers who learned English either before or after 21 years of age. Voice-onset-time is the measure between the release of the stop closure and the beginning of the next vowel. The Spanish speaker's VOTs from before 21 years old were more English-like than the speakers who were older than 21 years of age. These sorts of findings support the hypotheses of the Speech Learning Model (Flege, 1995), which suggest that differences between sounds in phonetic categories are responsible for foreign accents in late-learning students of language. There are a number of phonetic differences between Spanish and English. Gorman and Kester (2001) provide an extensive review of these differences. For instance, the nasals /m/ and /n/, when produced in clusters, are often interchanged by the Spanish-English bilingual, such as in the word "sometimes." In addition, there is a larger vowel inventory in English than in Spanish, leading to specific vowel sounds being difficult to produce by Spanish-English bilinguals. There are also differences in the place of articulation for certain phonemes. With [t] and [d], for example, these phonemes are pronounced on the alveolar ridge above the teeth in English, while in Spanish they are produced dentally, using the tip of the tongue on the back of the teeth.

In this study, I specifically focus on two features of Spanish-accented English. The first concerns the production of prevocalic stop consonants [p], [t], [k] and [b], [d], [g] in syllable initial position (Gorman & Kester, 2001). When native Spanish speakers produce English voiceless stops [p], [t], [k], they do not produce the accompanying aspiration which sometimes leads them to be perceived as the voiced /b/, /d/, and /g/ respectively by native English listeners. This occurs because the voice-onset-time of the

pre-vocalic, voiceless stop consonants is shorter when produced by a Spanish speaker than an English speaker (Lisker & Abramson, 1965). Voice-onset-time has been shown to be partially responsible for the perception of voicing in stop consonants. In this case, a shorter voice-onset-time results in the perception of the voiced counterpart. The second dimension of difference is the postvocalic stop consonant voicing (Morrison, 2002). Postvocalic stop consonants do not exist in Spanish, which leads to devoicing in the word-final stop consonants (Flege, Munro, & Skelton, 1992). For example, when [b], [d], [g] are produced in words like *mob*, *mad*, and *bag* these words are likely to be perceived as containing their respective voiceless counterparts (*mop*, *mat*, and *back*) respectively. That is, Spanish-accented post-vocalic stops are more likely to be perceived as unvoiced, rather than voiced by English listeners. These linguistic differences result in Spanish speakers having a distinct accent when producing English words that contain these stops (You & Kazemzadah, 2005).

In this study, I investigated how these differences affect the intelligibility of Spanish-accented English by using a word-recognition task. Native English listeners were presented with a token from either a native English speaker or Spanish-English bilingual. Then, they were presented with two words visually and asked to identify which of the two words it is or whether the correct choice is not present. Four factors were manipulated: The language background of the speaker (Native English versus Spanish-English bilingual), the voicing of the auditory word (voiced or voiceless), the vocalic position of the auditory word (pre- or post-vocalic), and the arrangement of the visual words. Using such a task permitted the evaluation of the key question of the study. I investigated whether the chosen phonetic dimensions reduce the intelligibility of the

accented speech relative to control, native speech. This study can provide further evidence suggesting that certain segmental differences are standalone in their contribution to accent perception. Furthermore, the results of this study could help identify specific differences in the perception of Spanish-English accented words. If these features result in lower accuracy, this could help identify an area which second language learners of English could focus on to reduce accentedness. However, if there are no differences in perception perhaps the features I am studying display more accentedness when interacting with suprasegmental features, or even other phonetic features.

Method

Participants

One native English speaker and one Spanish-English bilingual, native in Spanish, produced the stimuli. Each speaker's gender was female, as to avoid any gender effects. Initially, two speakers from each language background produced the tokens three times. Speakers also produced an English elicitation paragraph which contained difficult to pronounce English words for the purpose of determining the degree of accent the speaker exhibits. Recording was done in a sound-reduced booth using a microphone (SHURE SM-58) and the stimuli were recorded in Audacity® (Mazzoni, 2014). Then, for each language background the speaker's tokens were compared to the other speaker. Tokens that exemplified the essential differences between the two languages were used.

Nineteen adult listeners with self-reported normal hearing and eyesight participated in the experiment. Participants completed a demographic form to determine their language background, as well as their exposure to Spanish and Spanish-accented

English. Four participants were excluded from the study, due to substantial past experience with Spanish. Among the fifteen participants analyzed the age range was 19-46. Upon completion, participants were given either five dollars or course credits for their time.

Materials

The words used consisted of one-syllable, three phoneme CVC words (see appendix A). Each of the critical words were minimal pairs that differed on voicing. There were two locations of difference: pre-vocalic stop consonants and post-vocalic stop consonants. No word included more than one dimension. Filler words differed on either their nasality or manner but not on voicing. Words were chosen using the MRC Psycholinguistic Database (Wilson, 1987). The words were matched on their Thorndike-Lorge written frequency of use (Thorndike & Lorge, 1944) such that the overall average frequencies of the manipulated categories (i.e., voiced vs voiceless) were not statistically different.

Procedure

The participants were seated in front of a computer in a sound attenuated booth. They listened to sounds presented through headphones. On each trial minimal pairs were presented on the left and right of the screen. The participant was asked to identify which word was heard through the headphones. The participant would press the red button for the left word, blue for the right word, and green if the word was not present. The participant's accuracy and reaction time were measured.

Four factors were manipulated in this study. The first is the language background of the speaker presented to the participants, which has two levels: Native English and Spanish-English bilingual. The second is the voicing of the stop consonants presented to the participants, which has two levels: Voiced and unvoiced. Trials in which the auditory tokens consist of stop consonants that differ on voicing were the critical trials. The third manipulation is the position of the voicing, which has two levels: Pre-vocalic or post-vocalic. The fourth manipulation is the different visual alternatives presented to the participants. This factor has four levels: Target present-Competitor present (TPCP), Target present-Competitor absent (TPCA), Target absent-Competitor present (TACP), Target absent-Competitor absent (TACA). Target here refers to the auditory token presented to the participants and competitor refers to the voicing alternative to the target (e.g. if “pay” is produced, “pay” is the target and “bay” is the competitor). Table 1 below provides an example of the four levels of this factor for the target word “pay”.

There were also filler trials, in which the auditory tokens presented to the participants do not consist of stop consonants, but instead differ on some consonantal feature, such as nasals. There were two types of filler trials: Filler present and filler absent. In the filler present condition, the filler word is visually presented (e.g. if “may” is the filler, the choices will be “may” or “pay”). In the filler absent condition the filler is not visually presented (e.g. if “may” is the filler, the choices will be “say” or “pay”).

Results and Discussion

The English condition was included in order to ensure that there were differences between the Spanish and English speakers. There were two dependent variables in this study: Accuracy and reaction time. There were three independent variables manipulated

in this study: Voicing (voiced, voiceless), vocalic position (pre-, post-vocalic), and visual condition (TPCP, TPCA, TACP, TACA). A factorial ANOVA was used to analyze the data in the Spanish and English conditions. The English condition was analyzed to ensure that there was a difference between language groups and that there were no differences within the English condition on any condition. As expected, there was a significant difference between the two groups in accuracy, $F(1, 15) = 66.11, p < .01$. The mean accuracy was higher in the English condition ($M = 95.41, SD = 3.30$) than the Spanish condition ($M = 82.78, SD = 3.05$), which suggests that participants were worse when they heard the Spanish tokens as opposed to the English tokens (Figure 1A). In addition, there was a significant difference between the two groups' reaction time¹, $F(1, 13) = 20.23, p < .01$. The mean reaction time was lower in the English condition ($M = 1499.73, SD = 88.09$) than the Spanish condition ($M = 1880.29, SD = 125.75$), suggesting that participants responded more quickly when they heard the English tokens, as opposed to the Spanish tokens (Figure 2A). There were no significant differences in accuracy or reaction time in the English condition as a result of the manipulated factors, which was as expected.

The Spanish condition was the manipulation of interest, and, as such this condition was analyzed with increased scrutiny. There was a significant main effect of voicing, $F(1, 14) = 5.17, p < .05$, which was unexpected. Participants were more accurate in the voiced condition ($M = 90.55, SD = 2.87$) than the voiceless condition ($M = 87.64, SD = 3.04$). A main effect of voicing was not expected because the voicing of the tokens was not predicted to influence the perception or production of accented speech. This finding might suggest that whether the token is voiced or voiceless has an

¹ Only correct responses were analyzed in the reaction time data.

effect on the intelligibility of the speech. There was no significant main effect of vocalic position, $F < 1$ suggesting that whether the tokens differed in vocalic position did not influence accuracy, which was expected. Neither voicing nor vocalic position had overall effects on the accuracy of the perception of accented speech. Finally, there was no main effect of visual condition, $F(3, 12) = 1.12$, $p = .38$, which was not as predicted. The TPCP condition was expected to have worse results than the TACA condition because the task is more difficult in the TPCP condition. However, the visual condition had no influence on participants' accuracy. There was no significant interaction between vocalic position and visual condition, $F < 1$. As expected, whether the tokens were pre- or post-vocalic in any of the 4 visual conditions did not affect participants' accuracy. As predicted, there was a significant interaction between voicing and vocalic position, $F(3, 42) = 5.54$, $p < .01$, $\eta_p^2 = .56$. Participants were more accurate with voiced pre-vocalic words ($M = 93.88$, $SD = 3.28$) than with unvoiced pre-vocalic words ($M = 84.17$, $SD = 3.08$), and they were less accurate with voiced post-vocalic words ($M = 87.22$, $SD = 2.52$) than with unvoiced post-vocalic words ($M = 91.11$, $SD = 4.06$). The pattern in the pre-vocalic position is the opposite as the post-vocalic position (Figure 3A). This finding provides supporting evidence for the question of this study. In addition, there was a significant interaction between voicing and visual condition, $F(3, 42) = 5.41$, $p < .01$, $\eta_p^2 = .28$. This finding was unexpected. It implies that participants' performance was influenced differently by whether the token was voiced or not in each of the visual conditions. Participants were more accurate in the voiced TPCP and TPCA conditions than the voiceless counterparts. They were less accurate in the voiced TACP condition than the unvoiced TACP condition which suggests that performance was worse when the

tokens were voiced in the TACP condition (Figure 4A). There was also a significant interaction between voicing, vocalic position, and visual condition, $F(3, 42) = 5.72$, $p < .01$, $\eta_p^2 = .29$. The predicted interaction between voicing and vocalic position was influenced by the visual condition. As expected, for the voiced, pre-vocalic position, all four visual conditions resulted in higher accuracy than the voiceless, pre-vocalic position, and the voiceless post-vocalic position resulted in higher accuracy than the voiced, post-vocalic condition for all visual conditions.

For reaction time only the data in which the participants gave the correct response was analyzed. This is because participants might be slower to answer if they don't know the answer, thus the overall reaction time would be slower. I predicted that participants would have faster reaction time with the voiced pre-vocalic tokens than the voiceless pre-vocalic tokens and that participants have slower reaction time with the voiced post-vocalic tokens than the voiceless post-vocalic tokens. In addition, I also predicted that participants would react slower in the TPCP condition than in the TPCA and TACP conditions, and slower in those conditions than the TACA condition. As expected there were no significant main effects of voicing, $F < 1$. There was a significant main effect of vocalic position, $F(1, 13) = 13.07$, $p < .01$, $\eta_p^2 = .50$, which was unexpected. Participants were faster in the pre-vocalic condition ($M = 1573.90$, $SD = 88.92$) than the post-vocalic position (1806.12 , $SD = 119.00$). This might be because participants hear the pre-vocalic tokens before the post-vocalic tokens. However, this finding is not relevant to my question and will not be further discussed. There was not a significant main effect of visual condition, $F(3, 39) = 2.42$, $p = .08$, which was unexpected. I expected reaction time to be faster in the TPCP condition than the TACA condition. There was no

significant interaction between vocalic position and visual condition, $F(3, 39) = 1.50$, $p = .23$, which was expected. In addition, there was no significant interaction between voicing and vocalic position, $F < 1$, which was unexpected. I expected that the position of voicing would influence the participants' response time. There was, however, a significant interaction between voicing and visual condition, $F(3, 39) = 5.82$, $p < .01$, $\eta_p^2 = .31$. Participants were slower in the TPCA condition when the token was voiceless ($M = 1935.30$, $SD = 162.78$) than when it was voiced ($M = 1589.20$, $SD = 110.57$), and participants were slower in the TACP condition when the token was voiced ($M = 1892.82$, $SD = 166.35$) than when it was voiceless ($M = 1573.15$, $SD = 99.42$). Perhaps when the target was present it was easier to process voiced tokens than unvoiced tokens, but not in either of the extreme conditions. Finally, there was no significant three-way interaction, $F < 1$. Overall, the predictions were supported on the expected factors of accuracy, however there were some unexpected findings in reaction times.

The purpose of this study was to examine the extent that phonetic properties have on the perception of accent. When native Spanish speakers are learning English there are certain phonetic features of English that they have difficulty producing due to differences between the languages. I explored two specific features of Spanish-accented English: Pre-vocalic voiceless consonants sounding like the voiced counterpart and post-vocalic voiced consonants sounding like the voiceless counterpart. In conjunction with these dimensions, the task used in this study could separate each dimension individually and examine how strong of an effect specific phonetic properties have on accent perception.

I predicted that Spanish-English productions of pre-vocalic voiceless and post-vocalic voiced consonants would result in worse performance in the task than any of the

other tokens. The assumption is that worse accuracy in performing the task would represent that the participant perceived higher levels of accent. The results supported this prediction. This suggests that these dimensions might result in the perception of accented speech from a Spanish-English bilingual. Furthermore, this finding provides further evidence that phonetic features alone can be responsible for the perception of accent without the presence of suprasegmental features.

In addition to exploring the effect accent has on performance, I examined a task-related factor in the visual condition as well. Participants had to choose between two options presented on the computer screen, or select that the word was not present. The auditory token being presented was considered the target and its voicing counterpart was considered the competitor, and either one could be present or absent on the screen. This creates four visual conditions participants would be in: Target-present competitor-present (TPCP), target-absent competitor-absent (TPCA), target-absent competitor-present (TACP), or target-absent competitor-absent (TACA). I predicted that the TPCP condition would be the most difficult and result in the worst performance because both options presented would be valid. The TPCA and TACP would still provide a challenge, however the performance in this condition would be better than the TPCP condition. Finally, I predicted that the TACA condition would be the easier condition and would result in the best performance. I predicted that this would occur regardless of the auditory condition because it represents a difficulty in the task. The results did not support this prediction, however. There was no difference in performance between any of the four visual conditions. This implies that one level of this condition was not more difficult than another, regardless of other factors. However, visual condition did produce the expected

results in the context of the studied dimensions. The TPCP condition did produce worse results than the TACA condition when voicing and vocalic position were considered. These two findings suggest that the task was not more difficult, unless the auditory tokens were harder to understand. When the auditory tokens are easily discernable it does not seem to matter whether the target or competitor is present. The task was not more difficult if the tokens contain no features that would make them difficult to understand.

In addition to accuracy, I used reaction time as a measure of perceived accent such that faster reaction time might imply faster processing time. If the tokens were more difficult to understand because of the amount of accent that was exhibited then the reaction time would be slower because it would take longer to process. As such, I predicted that reaction time would be slower in both the pre-vocalic voiceless and post-vocalic voiced conditions, as opposed to their voicing competitors. This dependent variable produced some interesting and unexpected results. The predicted interaction did not occur, but instead, voicing had an influence on the reaction time in the TPCA and TACP conditions. Participants had slower reaction time in the TPCA condition if the token was voiceless rather than voiced, and the opposite was true in the TACP condition. This means that regardless of the vocalic position, voicing influences processing time. Finding an effect on reaction time but not accuracy suggests that this condition is not more difficult, but takes the participant longer to discern. Perhaps when both the voicing conditions are present participants have a visual comparison of the voicing alternative, which increases the speed of their choice. When neither choice is present it results in a similar effect. This might occur because participants do not have to waste processing time thinking about possible alternatives, and can use their visual perception to aid in

discerning the correct token. This was not the only interesting finding of reaction time, however. In the English speaker condition, vocalic position affected reaction time. This was unexpected, since this was the manipulation check. One possibility for this would be that participants were exposed to the pre-vocalic differences slightly before the post-vocalic differences, by nature of the position, and this resulted in slightly faster processing time. This finding is interesting, but not relevant to the study at hand so I will not explore its significance too deeply.

For the purpose of my research question, reaction time is only a secondary measure. Accuracy is more relevant in terms of answering the question because performance directly assesses how accented speech influences the perception of auditory tokens. One reason for this is that reaction time primarily provides insight into how individuals processed the tokens. Accuracy is the measure that most clearly indicates the direction of performance. For my question I will now focus on accuracy alone.

How is the predicted finding explicable from the context of the Speech Learning Model and the Perceptual Assimilation Model? The reason the phonetic dimensions I focused on resulted in accented speech is because those phonetic qualities do not exist in Spanish, but do in English. Thus, when learning English the perceived phonetic differences between the two phonemes are high for the learners. From the perspective of the Speech Learning Model, the English language learners do not have a mental category to “hold” the phonemes in. Thus the learners do not have a comparison which would allow them to produce the phonemes in a native-like way. In this study, this lack of a pre-existing category resulted in the perception of an accent, which resulted in worse performance from the participants. Similarly, the Perceptual Assimilation Model asserts

that this occurred because of differences in production of phonemes. Because the position of these particular dimensions do not exist, the means of articulating them are different. The co-articulation of the phones in Spanish-accented English is not done in a native-like manner. Both of these models assert that there is a difference in production and perception of the phonemes. The two dimensions I focused on resulted in worse performance, which supports the suggestions of these two models.

One of the limitations of this study is the number of speakers used. It is possible that the participants became attuned to each speaker because there was only one from each language background. Using another speaker would help account for any of this effect that might occur. In addition, in future studies I would like to include a Spanish-English listener condition to see if the opposite effects occur. This would also help transition to studying the effects of the Interlanguage Speech Intelligibility Benefit (Bent & Bradlow, 2003), which proposes that speakers of a foreign language can more easily understand that language when speaking to a non-native speaker of any language background, as opposed to a native speaker. In addition, having the same language background of a speaker further improves intelligibility. Specifically, I want to explore how phonetic features influence perception in the context of non-native L2 communication.

Finally, past research suggests that preceding vowel duration can influence the perception of voicing in post-vocalic stop consonants (Raphael, 1971). If the preceding vowel is longer, this can result in the perception of the voiced stop consonant, and if the vowel is shorter this can result in the perception of the voiceless stop consonant. This phenomenon could be interesting to explore in future studies because perhaps differences

in Spanish-English vowel length could interact with the differences in Spanish-English voice-onset-time.

In conclusion, this study further suggests that phonetic differences between languages can influence whether an accent is present in non-native speech. The findings in this study support assertions that the lack of a category for new phonemes could be responsible for accent formation, as well as the perception of accented speech.

References

- Abramson, A. S. & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. New York, NY: Haskins Laboratories.
- Bent, T., & Bradlow, A. (2003). The interlanguage speech intelligibility benefit. *Acoustical Society of America*, 114(3), 1600-1610.
- Bertran, A. P., (1999). Prosodic typology: On the dichotomy between stress-timed and syllable-timed languages. *Language Design* (2), 103-130.
- Best, C. T. (1995). A direct realist perspective on cross-language speech perception. *Speech perception and linguistic experience: Issues in cross-language research*, 171-204.
- Derwing, T. & Munro, M. (2003). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49, 285-310.
- Fernandez, E. M. (2005). The prosody produced by Spanish-English bilinguals: A preliminary investigation and implications for sentences processing. *Aberlin*, 4(1), 109-141.
- Flege, J. E. (1995). Second language speech learning theory, findings, and problems. *Speech Perception and Linguistic Experience*, 233-277.
- Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America*, 92 (1), 128-143.

- Gorman, K. B., Kester, E. S. (2001). Spanish Influenced English: typical phonological patterns in the English language learner. *Seminar offered by Bilinguistics at ASHA*.
- Kartushina, N. & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Front Psychology*, 5, 1246.
- Kondaurova, M. V. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, 124(6), 3959–3971.
- Mazzoni, D. (2014). Audacity® (Version 2.1.0) [Computer software]. Available from <http://audacity.sourceforge.net/download/>
- Morrison, G. S. (2002). Spanish listeners' use of vowel spectral properties as cues to post-vocalic consonant voicing in English. *The Journal of the Acoustical Society of America*, 112 (5), 2385
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42–46.
- Raphael, Lawrence, J. (1971). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America*, 51 (4), 1296-1303.
- Weinberger, Steven. (2015). *Speech Accent Archive*. George Mason University. Retrieved from <http://accent.gmu.edu>

Williams, D., & Escudero, P. (2014). Native and non-native speech perception. *Acoustics Australia*, 42(2), 79-84.

Wilson, Michael (1987, April 1). *MRC Psycholinguistic Database*. Retrieved from <http://websites.psychology.uwa.edu.au/school/MRCDatabase/mrc2.html>

Appendix A

Figure 1A

Main Effect of Accuracy in the Spanish and English Conditions

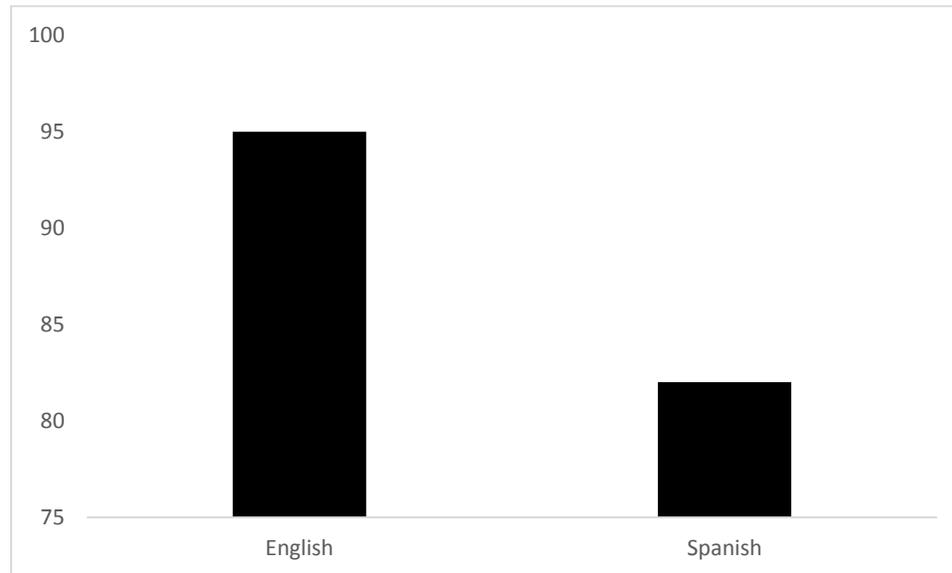


Figure 1A. This depicts the difference in accuracy between the Spanish and English condition.

Figure 2A

Main Effect of Accuracy in the Spanish and English Conditions

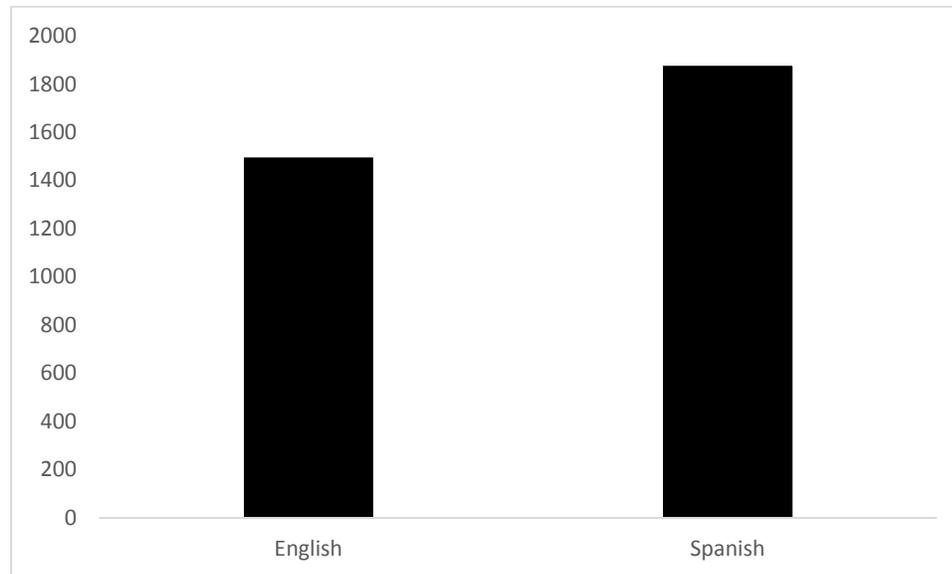


Figure 2A. This depicts the difference in reaction time between the Spanish and English condition.

Figure 3A

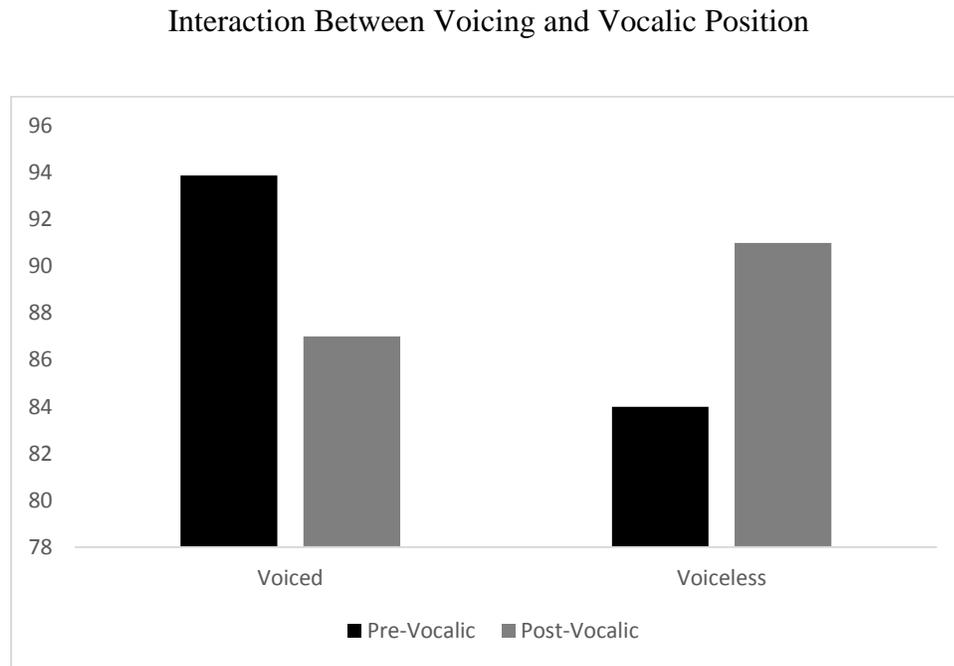


Figure 3A. This figure illustrates how the effect has switched from pre- to post-vocalic tokens.

Figure 4A

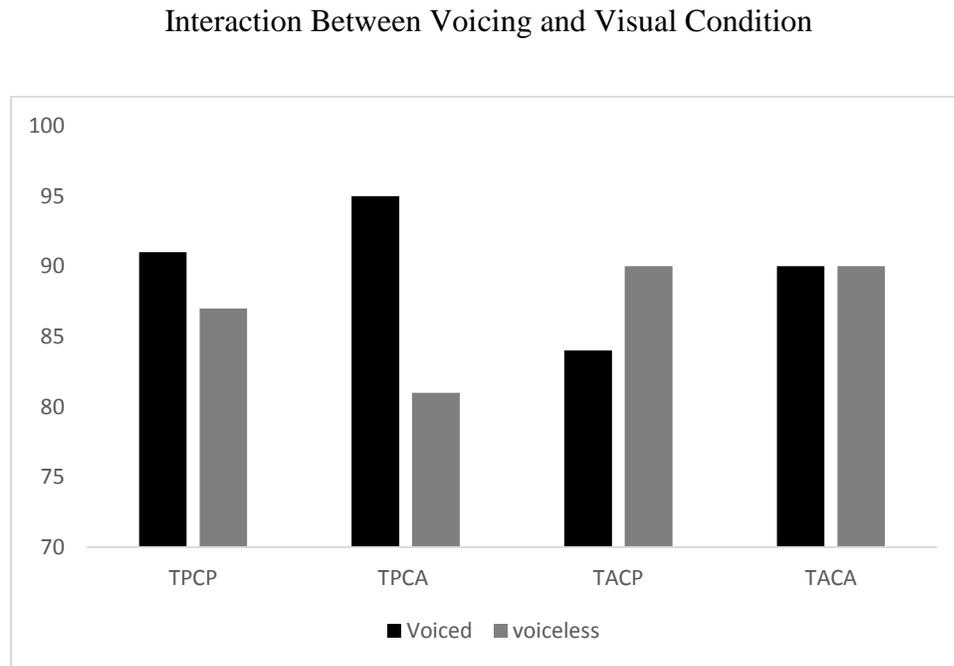


Figure 4A. This figure illustrates the unexpected finding that voicing is affected by the visual condition.

Table 1

Description of Visual Manipulation

Competitor Target	Present (CP)	Absent (CA)
Present (TP)	PAY - BAY	PAY-MAY
Absent (TA)	BAY- MAY	SAY-MAY

Table 2

Pre-Vocalic Word List

<u>Voiced</u>	<u>Unvoiced</u>	<u>Unrelated 1</u>	<u>Unrelated 2</u>
Bail	Pail	Sail	Mail
Deal	Teal	Seal	Meal
Gale	Kale	Sale	Male

Post-Vocalic Word List

<u>Voiced</u>	<u>Unvoiced</u>	<u>Unrelated 1</u>	<u>Unrelated 2</u>
Mob	Mop	Moss	Mom
Laid	Late	Lace	Lame
Hag	Hack	Has	Ham