

An Emerging Area in Second Language Phonology: The Perception of English Vowels by Adult Second Language Learners

Juri Matsubara¹

Introduction

Although the field of second language acquisition (SLA) has largely advanced over the last few decades, the area of second language (L2) phonology has not been a focus of many studies. Major (1998) found that “of the nearly 200 articles published in *Studies in SLA* [from 1988 to 1998], only about a dozen focused on phonetics and phonology” (p. 131). Since then, the area has largely expanded, as discussed in Eckman (2004), evolving from studies that cited first language (L1) influence as a main explanation for L2 phonology to research that has turned to universal principles driven by Universal Grammar (UG). Several crosslinguistic speech perception models have been developed as well, and many empirical studies have been undertaken to yield support. One of the most fertile domains for the investigation of L2 phonology has been the syllable (Eckman, 2004, p. 527). However, most of the studies have concentrated on the onset or the coda, while there is much yet to be explored for the acquisition of the nucleus of the syllable – the vowel. This paper is a literature review of the emerging area of L2 phonology, focusing on the perception of English vowels by adult L2 learners. First, the background will be presented by describing several theoretical models for L2 speech perception. A detailed description of empirical studies that seek to support these models will follow. Next,

¹ Julie Matsubara received her M.A. degree in Applied Linguistics from Teachers College. She is currently a lecturer at Kanda University of International Studies in Chiba, Japan. In addition to teaching in the English Language Institute, her duties involve materials development, curriculum development, and research. Her recent project explores how pronunciation activities can be incorporated into a reading course; her future endeavors include developing a curriculum for a pronunciation lab. Correspondence should be sent to: juliemats@gmail.com

this paper will explore whether or not these models are compatible with each other and if they could be consolidated into one. Finally, the future direction of L2 phonology will be discussed.

Theoretical Frameworks

The roots of L2 phonology can be traced to the contrastive analysis hypothesis (CAH) posed by Lado (1957), which postulated that comparing the L1 and the L2 of a particular language user will enable the prediction of difficulties that will be encountered in the acquisition of the L2. Lado stated that the “adult speaker of one language cannot easily hear language sounds other than those of his [or her] native language” (p. 11). In other words, the L2 phonemes will not be perceived correctly if a certain phonemic difference does not exist in the L1. The example Lado gave was of the Thai aspirated and unaspirated /p/. English speakers do not hear the phonemic difference in the two Thai /p/s, because these two sounds are allophones in English. In the case of vowels, one example is the English /i/-/I/ contrast that does not exist in Japanese, leading L1 Japanese listeners to not hear the phonemic difference between the two English vowels, according to the CAH.

Although the CAH may have been compelling and intuitive, it could only explain a portion of pronunciation errors made by L2 learners (Eckman, 2004). CA could not predict which particular L2 sounds would be difficult, and “never made it clear whether the difficulties encountered by L2 learners had a primarily motoric or perceptual basis” (Flege, 1992b, p. 567). Even though the CAH soon came into disfavor, Eckman (1987) noted that “rather than being abandoned altogether, the CAH should be revised to incorporate a notion of degree of difficulty which corresponds to the notion of typological markedness” (p. 68). However, this notion will not be dealt with in detail here because of the lack of empirical studies examining vowel perception under the framework of typological markedness.

Though the CAH fell out of favor among SLA researchers, several crosslinguistic speech perception models based on similarities and dissimilarities between the L1 and L2 emerged in the area of L2 phonology. One of the most prominent of these was the speech learning model (SLM) (Flege, 1992a), which postulated that phonetic categories that are needed to perceive L2 sounds rapidly and accurately in conversational speech can be established for *new*, but not *similar*, sounds for adult L2 learners (p. 162). This is because *similar* L2 sounds will be equated with existing L1 sounds, even if the sounds may differ acoustically and audibly, leading to equivalence classification (Flege, 1992b).

The implication is that if a certain portion of the phonetic space has not been used previously by the L1 vowel system, adult learners will be able to develop additional phonetic categories for new L2 sounds. Thus, the SLM leads to the prediction that although adult L2 learners may not perceive *new* L2 sounds authentically at first, they “will ultimately be more successful in [perceiving them] if they differ *substantially* from L1 sounds than if they differ *just a little*” (Flege, 1992a, p. 187). Though what may constitute a *new* or *similar* sound was not defined absolutely by Flege (1992a, 1992b), he suggested basing the distinction on “differences in the perceived phonetic distance between sounds in the L2 and those in the L1” (Flege, 1992b, p. 573). One method is to define a *new* sound as an L2 sound that was represented by an International Phonetic Alphabet (IPA) symbol that was not used in the L1, and which displayed audible or acoustic differences. A *similar* sound, on the other hand, while displaying audible or acoustic differences from the L1 sound also, would be represented by the same IPA symbol.

Another prominent crosslinguistic speech perception model was the perceptual assimilation model (PAM) (Best & Strange, 1992), which posed that adult L2 listeners perceptually assimilated L2 phones to L1 categories. In other words, L2 phones are perceived in

terms of their similarities and dissimilarities to L1 phones (Best & Strange, 1992, p. 306). These similarities would be based on temporal and spatial properties of the speech articulators, such as the lips, tongue body, and the glottis. According to Best and Strange, there are four patterns of assimilation: (a) two L2 phones are perceived as two L1 phones, (b) two L2 phones are assimilated equally well or poorly into one L1 phone, (c) two L2 phones are assimilated into one L1 phone unequally, resulting in category goodness difference, and (d) two L2 phones are non-assimilable to any L1 phonemes because of such a big difference. It is important to keep in mind that all of these patterns are based on the listener's perception of similarities, regardless of actual acoustic properties.

Based on the abovementioned patterns, Best and Strange (1992) made the following predictions about the discrimination of L2 phones by an adult L2 learner: the two L2 phones in pattern (b) would be the most difficult to discriminate, pattern (a) would lead to good discrimination, and patterns (c) and (d) would result in intermediate discrimination. The researchers noted that pattern (c) would be based on perceived phonetic similarity, whereas pattern (d) would depend on perceived acoustic similarity, since these phonemes would be heard as non-speech sounds. Best and Strange also discussed the effect of L2 experience on these patterns of perceptual assimilation: "Increased L2 experience may foster improved recognition of the *discrepancies* between the L1 and L2 phones" (p. 307). This could result in a decreased amount of assimilation of L2 phones to L1 categories, leading to the emergence of a new category within the listener's perceptual system.

The final theoretical model that will be discussed in this review was outlined in Prince and Smolensky (1993) as Optimality Theory (OT), which is an output-based theory, as opposed to input-based theories that were guided by UG. However, OT is still committed to UG, on

which any serious theory of phonology must rely, according to the authors. As opposed to the traditional view that constraints in a grammar are mutually consistent, OT argues that such constraints are highly conflicting and are not in accordance with each other concerning the well-formedness of a certain phonological structure. OT resolves this conflict by a strict domination hierarchy, by which the constraints are ranked, where any constraint that is higher on the hierarchy has priority over others which are lower. This hierarchy can determine which particular input is satisfactory within a certain perceptual system.

As summarized in this section, L2 phonology has moved beyond the CAH to be analyzed by speech perception models such as the SLM and PAM. A viable theory of OT has also emerged to provide an explanation for the difficulties adult learners may face when learning an L2. In the following section, various studies and their results will be presented for the purposes of providing support for these claims on adult speech perception.

Empirical Studies

The Nature of Contrastive Cues

Before delving into the respective framework-specific studies, it is important to explore the nature of contrastive cues that lead to differences between L1 and L2 vowel perception. Fox, Flege, and Munro (1995) investigated the nature of English vowel perception by L1 Spanish speakers, compared with the perception of Spanish vowels by L1 English speakers. The stimuli consisted of seven English and three Spanish vowels that were arranged in pairs. The participants, who were 30 monolingual English listeners and 30 L1 Spanish L2 English listeners, heard and rated the vowel pairs on a 1 (very similar) to 9 (very dissimilar) scale. The L1 Spanish listeners were found to use two dimensions of vowel perception, namely vowel height and distribution of the vowels in a two-dimensional perceptual plane. The English monolinguals, on

the other hand, utilized three dimensions – vowel height, a front-back distinction, and a central-noncentral distinction.

The researchers (Fox et al., 1995) next sought to find out whether or not L2 proficiency affected the perceptual dimensions used by the L1 Spanish listeners to identify English vowels. Proficiency level was determined by self-ratings and the experimenter's ratings of the subjects' ability to pronounce English. Proficient and non-proficient listeners alike only used the two dimensions of vowel height and distribution. In other words, learning English did not increase the participants' perceptual dimensionality. However, the greater the L1 Spanish subjects' English proficiency, the greater the reliance was on the distribution of the vowels in a two-dimensional perceptual plane. Therefore, although proficiency did not alter dimensionality, it did affect the nature of the participants' vowel perception.

Ingram and Park (1997) examined cross-language vowel perception by adult Japanese and Korean learners of English by presenting data concerning the effect of prior L1 phonological learning on the integration of temporal and vowel quality features in the perception of English vowels (Ingram & Park, 1997, p. 345). The 23 subjects were placed into subgroups according to how much English experience in the L2 environment they had, and there was also a control group of 8 L1 Australian English speakers. The participants were asked to perform a vowel identification task of standard Australian English by listening to vowel tokens produced by two male Australian speakers. After circling one of five choices on an answer sheet indicating which vowel they heard, they participated in another experiment where “perceptual judgments of the same five vowels were obtained from the standpoint of the listeners' native language” (Ingram & Park, 1997, p. 359). Subjects were given 7 seconds to transcribe an English item in Japanese

Kana or Korean Hangul, and then rated how close each English vowel item matched the Japanese or Korean vowel chosen in the transcription.

The researchers (Ingram & Park, 1997) found that the L1 Japanese listeners outperformed the L1 Korean listeners regardless of English experience. Moreover, L1 Japanese subjects were sensitive to speaker-related durations of English vowels, whereas the L1 Korean subjects utilized absolute duration cues more frequently. These results were attributed to “native-category transfer effects in the perception of non-native vowels” (Ingram & Park, 1997, p. 362), which can be explained by both the SLM and PAM. Another interesting observation was that there was a perceptual difference between older and younger Korean participants. This age effect was attributed to the recent phonological merger of Korean /e/ and /ɛ/ in the Korean language, especially for the younger generation in Seoul, Korea.

Flege, Bohn, and Jang (1997) also explored the effect of English experience on the perception of English vowels by adult learners, based on length of residence in the United States (US). The participants were 90 L1 German, Korean, Mandarin, and Spanish speakers who were asked to listen to synthetic vowel stimuli in a booth and push a button corresponding to the vowel they heard. The general finding was that more L2 experience led to more native-like vowel perception. L1 Korean listeners utilized more temporal cues overall, whereas Spanish participants used temporal or spectral cues for the /bit/-/bIt/ distinction, while using spectral cues for the /bæt/-/bet/ contrast. It was found that for the /bæt/-/bet/ difference, participants with more English experience placed greater weight on spectral cues than inexperienced subjects. Temporal cues were not used as much, and similar, but not as significant results, were found for the /bit/-/bIt/ contrast for Mandarin and Korean subjects as well. The researchers finally commented that

there was a crosslinguistic difference in how vowels in the L1 and English were perceived to be related.

Miranda and Strange (1989) addressed the importance of spectral, temporal, and dynamic cues for the perception of English vowels by adult L2 learners in comparison to native speakers of English. The study was motivated by the observation that the phonetic features that signal phonemic contrasts may be language-specific. The researchers attempted to find whether or not these phonetic features may differ between L1 and L2 English speakers. After electronically modifying American English vowels systematically by varying the amount of acoustic information available, the participants' patterns of vowel perception were observed according to target vowel information, dynamic information observed in formant transitions, and durational cues.

The results were remarkable for the researchers (Miranda & Strange, 1989), in that both L1 and L2 English listeners exhibited similar error patterns across the three types of phonetic features. This indicated that adult learners' perception of American English vowels did not differ much from that of the native speaker group. However, one subgroup of less-advanced L2 English participants performed poorly when the vowel stimuli were electronically modified, compared to when the vowels were unmodified. Therefore, it was concluded that all possible acoustic information may be needed for beginners to identify English vowels.

Escudero (2002) examined the perception of the Scottish English /i/-/I/ vowel contrast by L1 Spanish listeners. The premise for this study was that L1 speakers assign differing amounts of attention to various contrastive cues, an example being that L1 Scottish English speakers put more weight on spectral cues than durational cues, although both were used as principal cues. Escudero asked 30 L1 Spanish participants and 20 L1 speakers of Scottish English to listen to /i/

and /I/ sounds that were synthesized from naturally produced vowels of two L1 Scottish English speakers. Two tasks were administered, namely (a) a discrimination same/different test, where the subjects listened to two stimuli and indicated whether they were the same or different, and (b) an identification test, where upon listening to the stimulus, the participants pressed a button corresponding to a picture of a *sheep* or a *ship*, which the subjects were asked to identify beforehand. Pictures were used to avoid orthographic effects that could be brought on by differences in spelling.

The results of the study (Escudero, 2002) were consistent with other crosslinguistic perception studies. The L1 Scottish participants, as mentioned before, relied more on spectral than durational information. In contrast, the L1 Spanish group generally relied more on durational than spectral cues, but there were three main patterns that could be observed within the Spanish group: (a) One subgroup displayed high reliance on duration and negative reliance on spectral information, (b) another group relied highly on duration while displaying low reliance on spectral cues, and (c) the last group was similar to the L1 Scottish group in that they primarily used spectral cues to perceive the /i-/I/ contrast.

Support for the SLM

Although all three of the aforementioned theoretical models have been investigated extensively through L2 phonology studies, vowel perception studies are still limited. However, Flege's (1992a, 1992b) SLM model was found to be the basis of the following studies. Flege, Munro, and Fox (1994) examined auditory and categorical effects that could be observed for L2 vowel perception. The study was motivated by the belief that there exists a universal, sensory-based component in vowel perception across languages. The research participants were 60 paid listeners, of which 30 were American English monolinguals, and the remaining were 30 L1

Spanish L2 English speakers. The L1 Spanish subjects were placed into subgroups according to their English experience, which was based on how long they had lived in the US. In a sound booth, the participants were asked to rate 405 vowel pairs on how similar they perceived the vowels to be, on a scale from 1 (very similar) to 9 (very dissimilar). It was found that the amount of L2 experience did not have a significant effect on the perceived dissimilarity of vowel pairs, contrary to the researchers' expectations.

Flege et al. (1994) conducted a second experiment, in which the same participants performed an oddity task, where they listened to 108 vowel triads and identified the odd item out. The participants were to circle the item number that differed from the others, or circle 'N' to indicate that all the items were the same vowel. The subjects were once again divided into an English monolingual group, an L1 Spanish less English experience group (mean length of residence 1.8 years), and an L1 Spanish more English experience group (mean length of residence 7.0 years). The effect of group was not found to be significant for nonadjacent triads, where the odd vowel was in a different vowel space than the others, and was hence a *new* vowel for the L1 Spanish speakers. However, monolingual English listeners outperformed L1 Spanish listeners regardless of English experience for adjacent triads, where the odd item out was a vowel in the same vowel space as the others, or in other words, a *similar* vowel for the L1 Spanish speakers. The researchers concluded that these results provided support for the SLM, in that adult language learners would have difficulty discriminating L2 vowels that were *similar* to L1 vowels. However, the lack of a significant English experience effect led Flege et al. (1994) to ponder whether the age of arrival (AOA) in an L2 environment or the amount of L1 language usage would have a significant effect.

Therefore, Flege, MacKay, and Meador (1999) placed 72 L1 Italian participants into four subgroups based on their AOA in Canada and self-reported continued use of Italian. Eighteen monolingual English speakers also participated in the study as a control group. The study, similar to Flege et al. (1994), examined adult L2 vowel perception by means of a categorical discrimination test. There were 11 vowel contrasts, four being English-English vowels, four English-Italian, and three Italian-Italian. The stimuli were spoken by native female speakers of the respective languages, and the stimuli were derived from words and nonwords, as opposed to synthetic tokens used in some perception studies (e.g., Flege et al., 1997). Temporal differences in vowels were reduced to ensure that subjects would not discriminate vowels based solely on duration.

Similar to Flege et al. (1994), upon hearing the vowel triads, the participants in Flege et al. (1999) clicked button 1, 2, or 3 if one of the stimuli was different from the others, and the 'no' button if all were the same vowel. The researchers found that the subjects' scores decreased as the AOA increased, and that the main effect of vowel contrast was significant. For the Italian-Italian vowel contrast, there was no group effect. However, subjects who had arrived in Canada *early* (mean age of 7 years), regardless of their continued use of Italian, performed as well as the English monolingual speakers in discriminating between English-English vowels and English-Italian vowels. On the other hand, subjects in the *late* (mean age of 19 years) AOA group obtained lower discrimination scores, leading the researchers to conclude that the age at which L2 learning in the L2 environment begins has a profound effect on adult L2 vowel perception, rendering support for the SLM.

Although Flege et al. (1999) found no significant effect for amount of L1 usage, Flege and MacKay (2004) found differing results upon also examining the perception of English

vowels by L1 Italian adults who were placed into subgroups based on AOA in Canada and amount of L1 usage. First, difficult English vowel contrasts were identified by asking Italian students who had only been living in Canada for 3 months to perform a vowel identification task. The results for the /I/-/ε/ contrast suggested that “nonnative discrimination of L2 vowels may depend on more than just cross-language patterns of perceptual assimilation” (Flege & MacKay, 2004, p. 13). This was in reference to PAM and the four patterns of assimilation, and although Best and Strange (1992) only commented on the possibility of L2 experience resulting in a new perceptual category, extending this possibility to other factors, such as L1 usage, is quite plausible.

The researchers (Flege & MacKay, 2004) found that L1 Italian speakers who arrived early in Canada and also had low usage of continued Italian obtained similar vowel identification scores to native English speakers. However, this was not the case for participants who arrived early and reported higher usage of Italian. Moreover, L1 Italian speakers who arrived later and used Italian more obtained significantly lower scores than those who used Italian less. These results were considered support for the SLM, but more importantly, the implication was that L2 phonology needed to rethink the overt focus on the starting age of L2 learning. Although age was an important factor, the amount of L1 use also made a difference.

The last study to be explicated in this section is a longitudinal study (Morrison, 2002) on the perception of the Canadian English vowels /i/ and /I/ by L1 Japanese and Spanish listeners. Although both Japanese and Spanish have a five-vowel system, only Japanese has a phonemic contrast between long and short vowels. Therefore, Morrison hypothesized that according to the SLM, vowel perception by native speakers of these languages would differ: Canadian /i/ and /I/ would be perceived respectively as Japanese /i:/ and /i/, leading the perception of these vowels to

not achieve native-like quality; on the other hand, English /i/ would be perceived as more similar to Spanish /i/, and a new category would be formed for English /I/, leading to the eventual distinction of these vowels to better match English listeners.

Morrison (2002) collected data from 7 L1 Japanese and 5 L1 Mexican Spanish participants, who were undergraduate students studying abroad in Canada, on two occasions, one month and six months after their arrival in Canada. Data was also collected from a control group of 7 Canadian English participants. The stimuli was composed of English words in sentences produced by a male monolingual English speaker, the vowels /i/ and /I/ in a stressed position followed by a voiced or voiceless stop. Using the stimuli, Morrison created a multidimensional continuum along vowel spectra, vowel duration, stop closure duration, and speaking rate. The subjects were asked to listen to 350 stimuli and click on one of four pictures representing the word, or “X” to signify they didn’t hear any of the words. Pictures were once again used to avoid orthographic effects.

After analyzing the data quantitatively and determining the categorical boundary between /i/ and /I/ for the control group, Morrison (2002) found that for the L1 Japanese speakers, duration properties were the primary cue for vowel perception, and this did not change significantly after six months of English experience. However, although L1 Spanish speakers also only employed duration cues initially, they developed a categorical perception of English vowels similar to native English speakers after six months. This led Morrison to conclude that because of the perception that English /i/ and /I/ were similar to Japanese /i:/ and /i/, L1 Japanese speakers ultimately did not achieve native-like categorical perception. However, because L1 Spanish speakers created a new category for the English vowel /I/, they were able to achieve categorical vowel perception.

Support for the PAM

Although many L2 perception studies cite the PAM, there are surprisingly few empirical vowel studies that yield direct support for it. In fact, the one study discussed here was based on the SLM. Flege and Bohn (1989) examined the perception of English vowels by L1 Spanish speakers. The participants were to listen to the words *beat*, *bit*, *bet*, and *bat*, with the respective English vowels of /i, I, ε, æ/, and answer whether or not they heard one of the Spanish vowels of /i,e,a,o,u/. If they did not hear any Spanish vowel, they were to respond ‘none.’ Generally, the subjects answered Spanish /i/ for English /i/ and /I/, Spanish /e/ for English /ε/, and Spanish /a/ for English /æ/, supporting PAM. However, participants who could speak English answered ‘none’ significantly more often than those who could not, indicating that they may have started creating phonetic categories in their perceptual vowel inventory.

In a follow-up experiment, the researchers (Flege & Bohn, 1989) asked the same participants to listen to vowels that varied on a continuum based on spectral quality and vowel length. For example, one continuum ranged from /bet/ to /bæt/, while another ranged from /bit/ to /bIt/. For the /bet/-to-/bæt/ continuum, the L1 Spanish subjects showed a clear perceptual crossover from /ε/ to /æ/. However, for the /bit/-to-/bIt/ continuum, only 30% of the participants were able to detect the distinction as the spectral quality and duration of the vowel shifted from one end to the other. It was speculated that this occurred due to the endpoints of the /bet/-to-/bæt/ continuum being identified with the two respective Spanish vowels of /e/ and /a/, whereas the endpoints of the /bit/-to-/bIt/ continuum were identified with the single Spanish vowel of /i/.

A third experiment was undertaken to control the effects of vowel duration, but the patterns of perception for the continua did not change. This led Flege and Bohn (1989) to conclude that although /ε/ and /æ/ were *new* vowels for the L1 Spanish speakers, it did not seem

that the participants had established new phonetic categories for them. However, it should be noted that with more experience in English, such new categories may form. Also, the results are in line with Best and Strange (1992), since PAM predicts that perceiving the distinction between two L2 vowels that have been assimilated to two L1 vowels would be easier than perceiving the distinction between two L2 vowels that are assimilated into one L1 vowel.

Support for OT

Once again, there have been many L2 phonology studies based on OT, yet vowel perception studies are sparse. However, Escudero and Boersma (2004) performed a very compelling study that was based on OT by examining the acquisition of English /i/ and /I/, which differ in vowel height and length. The study intended to build on research (Escudero, 2002) that had shown that adult learners “may weigh the cues to phonological contrasts differently from native speakers of the L2” (Escudero & Boersma, 2004, p. 552) in vowel perception.

Specifically, L1 Spanish listeners are not sensitive to duration, and therefore have difficulty with English vowels that differ both in quality and length. The 30 L1 Spanish subjects ranged in age from 18 to 58 years old, and their L2 was either Scottish English or Southern British English, which they started learning after the age of 12. A control group of 20 Scottish Standard English and 21 Southern British English speakers also participated. The stimuli utilized in the study were 37 isolated synthetic /i/ and /I/ vowels, and a forced identification task was employed, where subjects listened to stimuli and then pressed a button that corresponded to what they heard. The buttons did not say /i, I/, but instead had a picture of a *ship* and a *sheep*, to avoid orthographic effects on the responses.

The subjects’ responses were analyzed by the researchers (Escudero & Boersma, 2004) based on reliance ratios, which displayed the differing degrees of reliance on duration versus

spectrum when perceiving /i/ and /I/. The L1 Scottish listeners had a clear preference for spectral cues, while more than half of the L1 Southern English listeners used a combination of cues. The L1 Spanish listeners were first divided into a beginners group and an advanced group based on their answers on a language background questionnaire. The beginners group was found to utilize a perceptual strategy based on the L1 Spanish /e/ and /i/. The advanced group was much more complex: Depending on their background, the participants were found to utilize only spectral information or durational information. For example, students with a higher education used only durational information, students who had spent more time in Scotland or Zimbabwe used only spectral information, and students who had spent more time in England used only durational information.

Based on the results, the researchers (Escudero & Boersma, 2004) came to the conclusion that there was a correlation between L2 perception and the target dialect, in this case Scottish English versus Southern British English. More than half of the L1 Spanish listeners relied mainly on duration, yet when the production environment consisted of Scottish English, the listeners depended mainly on spectral cues. How can this be explained? Escudero and Boersma found the solution in OT based on first formant frequency (F1) constraints and duration constraints, for example ““an F1 of 260 Hz should not be perceived as /I/’ and... ‘a duration of 50 ms should not be perceived as /i/’” (p. 565). These constraints establish an optimized vowel categorization, which results in L1 Spanish L2 Scottish English speakers forming two perceptual categories for /i/ and /I/, whereas L1 Spanish L2 Southern British English speakers formed a single category, forcing these listeners to split the L1 Spanish /i/ into two new vowels or form a new feature contrast of duration.

As can be seen by the empirical studies discussed, L2 vowel perception has been considered from the viewpoint of several theoretical frameworks based on the groundwork that listeners employ various contrastive cues when perceiving English vowels. The following section will consider whether or not these frameworks are compatible with each other, and how they might be consolidated into one.

Compatibility of the Theoretical Frameworks

The first thing to consider when discussing the compatibility of the SLM, PAM, and OT is that the three frameworks do not have a common starting point. The SLM and PAM analyze adult L2 vowel perception based on the similarities and dissimilarities between the L1 and L2, and therefore are inherently connected to the CAH. However, OT seems to radically differ, because it is constraint-based, ultimately leading back to UG. Therefore, a consolidation of the three seems unlikely. However, when examined closely, it appears the SLM, PAM, and OT can all be applied to vowel perception, only at different stages. For example, the PAM can be used to analyze the initial stage of L2 perception, the SLM for the ultimate acquisition of L2 vowels, and OT for the development between the initial and ultimate stages.

The best study to model this consolidation seems to be on Morrison (2002), since it is the only longitudinal attempt to observe L2 vowel perception. At the initial stage of vowel perception, the perceptual assimilation patterns between L1 Japanese and Spanish listeners would differ due to Japanese, but not Spanish, having a phonemic contrast between long and short vowels. As explained above, the Japanese listeners would follow pattern (a) of the PAM, in which the Canadian English /i/ and /I/ would be assimilated into the Japanese /i:/ and /i/, leading to good discrimination. The Spanish listeners, on the other hand, would follow pattern (c), in which the Canadian English /i/ and /I/ would be assimilated into the Spanish /i/ unequally,

resulting in category goodness difference and intermediate discrimination. Because Morrison (2002) did not utilize a discrimination test using natural stimuli, it is difficult to validate the PAM with the study's results for when the participants had been in Canada for 1 month. Nevertheless, this discussion shows that the PAM can be applied to Morrison's study.

The SLM, in turn, would address the ultimate acquisition of L2 vowels based on their relationship with the L1 categories. Because the Canadian /i/ and /I/ would be perceived as *similar* to Japanese /i:/ and /i/ respectively, the perception of these vowels would be predicted not to achieve native-like quality; Spanish /I/, on the other hand, would be perceived as *new*, leading to the prediction that ultimate distinction of these vowels would better match English listeners. This is in fact what Morrison set out to investigate, and found positive results. Therefore, the PAM and SLM are shown to be compatible with each other.

The final question is whether OT can be included in this picture. "Neither the SLM nor the PAM...is currently able to give an accurate and complete developmental account of L2 speech perception" (Escudero & Boersma, 2004, p. 564). In other words, although the PAM could be utilized for initial vowel perception and the SLM for ultimate vowel perception, they do not explain how an adult L2 learner develops from one stage to the other. The reason why OT would be able to provide this explanation is due to the strict domination hierarchy, by which the constraints are ranked. As a learner moves between stages, the constraints on the hierarchy become re-ranked, thereby leading to a change in the nature of his or her vowel perception. This re-ranking was not examined by Morrison (2002) because OT was not considered in the study. However, the study could gain even more credibility by examining the developmental stages of the L1 Japanese and Spanish students using a constraint-based analysis.

Although it seems that no studies have consolidated the three theoretical models into one for the purposes of investigating L2 vowel perception, the present discussion points to a potential compatibility of the frameworks. Although the theory behind the models may be drastically different (i.e., CAH vs. UG), it seems of utmost importance for L2 researchers to acknowledge the potential benefits each framework can achieve by filling in the gaps in the research.

The Future of L2 Vowel Perception

Based on the empirical studies that have been undertaken to date, it seems L2 vowel perception research can expand in many directions. First, the only variables that have been observed in terms of having an effect on vowel perception are: (a) amount of English experience or length of residence in an English-speaking country, (b) L2 proficiency, (c) AOA, and (d) amount of L1 use. Flege and MacKay (2004) stressed the need of additional research to explore the effect of differing amounts of native speaker input on L2 vowel perception. Amount of L2 use and learner awareness are also potential factors.

Another direction of expansion concerns the L1 to be analyzed for English vowel perception. German, Italian, Japanese, Korean, Mandarin, and Spanish have been examined in the studies above, but the majority have concentrated on Japanese and Spanish. Therefore, there is much work to be done to test the theoretical frameworks for numerous other languages.

Regarding the nature of experimental design, there is much room for development. As seen in this review, only one study (Morrison, 2002) was longitudinal. This is alarming, especially since the three theoretical frameworks focus on different stages in L2 vowel perception. Also, more research on the L2 perception of English back vowels is needed, given that the majority of studies above focused on the /i/-/I/ contrast. Finally, the nature of the stimuli (synthesized vs. words vs. nonwords) affecting vowel perception is one aspect to be considered.

Because L2 vowel perception is such a great concern for L2 learners, it seems that another natural direction for future research is to examine the effects of various teaching techniques and classroom environments on vowel perception. Of particular intrigue is the English as a foreign language environment where multiple dialects of English are provided in the input. Although effects of training in a laboratory setting have been studied (Nishi & Kewley-Port, 2005; Sperbeck, 2005), the effects of classroom learning and awareness-raising are still yet to be explored. Also, an account of how we may apply the current L2 vowel perception knowledge to the teaching of vowel perception is a ripe area to cultivate.

Conclusion

Although L2 phonology has come a long way the last couple of decades, adult L2 English vowel perception research has only started to emerge. Compelling research based on the SLM, PAM, and OT have been undertaken. However, in order for this area to continue expanding, not only a strong understanding of the theoretical frameworks is necessary, but also an open mind to the possible compatibility among the models is also pertinent. The future of L2 vowel perception research lies in explicating the effects of unexplored factors such as learner awareness, other L1s, and most importantly for the L2 educator, teaching techniques.

References

- Best, C., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants, *Journal of Phonetics*, 20, 305-331.
- Eckman, F. R. (1987). Markedness and the contrastive analysis hypothesis. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 55-69). Cambridge, MA: Newbury House.
- Eckman, F. R. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition*, 26, 513-549.
- Escudero, P. (2002). The perception of English vowel contrasts: Acoustic cue reliance in the development of new contrasts. In J. Leather & A. James (Eds.), *New sounds 2000: Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech* (pp. 122-131). Klagenfurt, Austria: University of Klagenfurt.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551-585.
- Flege, J. (1992a). The intelligibility of English vowels spoken by British and Dutch talkers. In R. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 157-232). Amsterdam: John Benjamins.
- Flege, J. (1992b). Speech learning in a second language. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and application* (pp. 565-604). Timonium, MD: York Press.
- Flege, J., & Bohn, O.-S. (1989). The perception of English vowels by native Spanish speakers. *Journal of the Acoustical Society of America, Suppl. 1.*, 85, S85(A).

- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.
- Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34.
- Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106, 2973-2987.
- Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95, 3623- 3641.
- Fox, R., Flege, J., & Munro, M. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America*, 97, 2540-2551.
- Ingram, J. C. L., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, 25, 343-370.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.
- Major, R. C. (1998). Interlanguage phonetics and phonology: An introduction. *Studies in Second Language Acquisition*, 20, 131-137.
- Miranda, S., & Strange, W. (1989). The role of spectral, temporal, and dynamic cues in the perception of English vowels by native and non-native speakers. *Journal of the Acoustical Society of America*, 85, S53(A).
- Morrison, G. S. (2002). Perception of English /i/ and /I/ by Japanese and Spanish listeners: Longitudinal results. In G. S. Morrison & L. Zsoldos (Eds.), *Proceedings of the North West Linguistics Conference 2002* (pp. 29-48). Burnaby, Canada: Simon Fraser University Linguistics Graduate Student Association.

Nishi, K., & Kewley-Port, D. (2005, May). *Training Japanese listeners to identify American English vowels*. Poster session presented at the 149th meeting of the Acoustical Society of America, Vancouver, Canada.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Tech. Rep. No. 2). New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Sperbeck, M. (2005, May). *Training Japanese L2 learners to perceive difficult American vowel contrasts*. Poster session presented at the 149th meeting of the Acoustical Society of America, Vancouver, Canada.