Methods for Researching Second Language Phonological and Phonetic Acquisition

Christine Shea
University of Iowa

Abstract

In this commentary I provide a brief overview of selected research areas and methodology used to study adult L2 phonology and phonetics. I focus the discussion on studies that include Spanish as either the first or target language.

1. Introduction

Speech perception and production are language-specific in nature. Extensive research has demonstrated that second language learners do not perceive or produce target language sounds in exactly the same way as native speakers (see Bohn & Munro 2007 and the articles therein) – experience with a native language influences the perception and production of sounds that differ either phonetically and/or phonologically from its sound inventory. Nonetheless, acquisition of a non-native sound system can occur, even if the final outcome is not identical to that of native speakers of the target language (Abrahamsson & Hyltenstam 2009). In this commentary, I focus on adult L2 acquisition of phonetics and phonology, concentrating on studies related to Spanish, either as a native language or target language and present a selection of studies that demonstrate the variety of research methods employed in the field. I consider adult L2 learners as those who began the acquisition of their second language after a putative Critical Period (after approximately 12 years of age).[1]

2. Methodological issues related to the learner

2.1 Introduction

Factors related to the learner, such as proficiency, age of acquisition, language use and context of learning, may be used as either initial grouping factors or independent variables themselves, depending upon the type of study being designed. In what follows, I discuss age of acquisition and context of learning as two crucial determinants of phonological acquisition in adult second language learners.

Studies in Hispanic and Lusophone Linguistics
Volume 6, Issue 1
Spring 2013

2.2 Age of arrival (AoA) and context of learning

Proponents of the Critical Period Hypothesis as it applies to second language learning maintain that native-like attainment by late learners of a second language is almost impossible. Due to loss of neural plasticity or a type of neurofunctional reorganization, complete mastery of an L2 is no longer possible if L2 learning begins after the age of 12-15 years (e.g. DeKeyser, Alfi-Shabtay & Ravid 2010, Lennenberg 1967). In the acquisition of second language phonetics and phonology, the effects of age of arrival in the L2 context (AoA) is one of the factors that is included in almost every study, either forming a part of the hypotheses to be tested or as a factor for participant grouping.

The age at which learners begin to acquire the sound system of their second language can also interact with the context of their learning – that is, whether the learner is acquiring the target language in an immersion context, a classroom context or a study-abroad context. Closely related to this is the notion of language 'use.' This refers to the amount and type of interaction the subject has in the target language. While it is entirely possible that an individual lives in a context where the target language is spoken (and began to acquire it at a relatively early post-PC age), if the target language is not used on a daily basis, proficiency may not be as great as someone whose AoA is higher but has greater opportunities to use the target language. Indeed, research has shown that earlier acquisition does not always lead to more native-like proficiency in terms of the acquisition of second language speech if the younger age does not coincide with rich and authentic native-speaker input and abundant opportunities to interact in the target language (Flege & Liu 2001). Thus, progress in learning an L2 is dependent upon both the quantity and the quality of the L2 input learners receive (Flege 2002).

In terms of methodology, the variables mentioned in this section can form the grouping variable for any type of experimental study. It is vital that they be clearly defined at the initial stages of the research project in order to account for the results in a meaningful fashion.

2.3 Measuring proficiency

Typically, studies of L2 production rely upon either spectrographic analysis to compare the learners' speech with a native-speaker baseline (see Simonet 2011, 2012) or, alternatively, rely upon native speaker judgments (Moyer 2007). Studies that rely upon listener evaluations are necessarily more holistic in nature and given this, the selection of raters is an important decision since raters with more experience will tend to listen for different things than those with less formal experience. In terms of the number of raters, 2-3 are sufficient to obtain reliable results from an inter-rater reliability measure such as Cohen's Kappa (see Moyer 1999, 2004, 2007). Rating scales themselves typically range from 1 to 9 points

(Flege, Yeni-Komshian & Liu 1999) with 5 to 6 being the most common (Moyer 2007). Studies using these more holistic measurements have found that indeed, there are post-CP learners who do achieve native-like pronunciation ratings (Dutch-French bilinguals, Bongaerts 1999) when compared to monolingual native speakers and less proficient L2 learners, although work by Hyltenstam & Abrahamsson (2009) has shown that learners whose speech samples were rated as native-like (length of residence >25yrs) were no longer rated as such after being subjected to more rigorous testing.

For research on speech perception and production, measuring proficiency can be difficult since fluency or morphosyntactic knowledge does not always directly correlate with phonological or perceptual ability (however this may be defined). In studies looking at morphosyntax, grammatical knowledge can be more directly measured via grammaticality judgment tasks, cloze tasks and error detection tasks. In speech learning, however, grammatical knowledge may not reflect directly onto phonological competence or phonetic knowledge (see Tremblay 2009, for a study that correlates the two using a global pronunciation proficiency scale). Recent work based upon tenets of the PAM-L2 (see below) by Bundgaard-Neilsen, Best & Michael Tyler (2011) suggests that, similar to children acquiring a first language (Pierrehumbert 2003), adult second language learners with larger vocabularies may assimilate target language sound categories in a more native-like fashion than those with smaller vocabularies. This intriguing and promising approach may shed light on how to conceptualize second language phonological proficiency.

3. Models

3.1 Introduction

The acquisition of a second language sound system can be examined from a number of different perspectives, each with distinct theoretical and methodological considerations. For example, researchers have examined the factors that play the strongest role in the detection of a foreign accent, which involves native listeners rating the speech of non-natives (see above and work by Derwing, Munro and colleagues). There is also a large body of research on how L2 learners acquire the suprasegmental aspects of speech such as stress (see work by Archibald 1992, 1993 and Guion, Harada & Clark 2004) and intonation (Henriksen, Geeslin & Willis 2011). Work by Davidson (2006) and Carlisle (1998, 1999) examines the acquisition of non-native phonotactic production. On the more phonetically-oriented side, research has been driven by two highly influential models of second language speech learning, the Speech Learning Model (SLM) (Flege 1995) and the Perceptual Assimilation Model (PAM) (Best 1995). Both models center on how sounds from the native language are assimilated into the categories (phonetic in the case of the

SLM; articulatory-based phonetic/phonological in the case of the PAM) of the native language.

The SLM focuses on the perception and production of L2 sounds. According to this model, L2 phonetic acquisition depends upon the way in which target language sounds are assimilated into the native language system. This, in turn, is based upon acoustic similarity. When the target language sound is close to the native language sound, the learner will be faced with a case of *equivalence* classification and the creation of new L2 categories will be impeded. For Flege, listeners perceive and acquire phonetic categories, not phonemes, and L2 phonetic categories may be classified as 'identical,' 'similar' or 'new' in terms of the first language sound inventory (Flege 1995 *inter alia*). Crucially, the mechanisms and processes used in learning the L1 system remain intact over the lifespan and the native system can evolve to include all sounds that share the same equivalence classification. Flege argues that the equivalence classification mechanism may lead to first language interference in the acquisition of the second: L2 listeners may categorize a target language sound as similar to an existing L1 sound and remain blocked from creating a new category.

The PAM has traditionally focused on naïve listeners but this focus has shifted to L2 listeners with the PAM-L2 model (Best & Tyler 2007). The PAM puts forth a taxonomy of how listeners are likely to assimilate new sounds, based upon native categories – whether 'good,' 'poor,' 'equivalent' or 'unclassifiable.' In a study testing PAM hypotheses, the methodology will typically include some sort of rating task or identification task (see below).

When investigating the production and perception of second language segments, researchers must have a clear idea of the relationship that exists in each language between the sounds under investigation at the phonological, phonetic and orthographic levels. For example, sounds may be acoustically similar (in terms of formant values or acoustic parameters) but represented by different orthographic symbols, a fact that will necessarily influence perception and production in literate learners. For example, the sounds represented by the letters 'd' and 'th' in English correspond to /d̪/ and the allophone [ð̞] in Spanish. Thus, in these cases, the same orthographic symbol corresponds to different sounds in each language and, most crucially, sounds that are phonemic *and* phonetic in Spanish but only phonemic in English. In order to examine the relevance of this difference in levels of representation, Boomershine, Hall, Hume & Johnson (2008) investigated the perceptual similarity of [d], [ɾ], and [ð] by English and Spanish listeners. In English [d] and [ð] are phonemic and [ɾ] is allophonic while in Spanish [d] and [ð] share an allophonic relationship and [ɾ] is its own phone. In a similarity rating task, listeners rated the sounds that are allophones in their respective native languages as more similar than the phonemically contrasting sounds. This finding of subjective perceptual similarity was echoed in the results of a speeded AX discrimination task where listeners had increased latency in responding "different" to pairs of VCV

sequences with consonantal allophones (e.g. English listeners responded more slowly to the pair [ada] and [aɾa]) than between "different" pairs of a phoneme versus an allophone (e.g.[ada] vs. [aða]). The results from this study show how important it is to consider the levels of representation corresponding to the second language sounds when determining how these sounds will be assimilated by non-native speakers.

3.2 Speech perception research: Methods

In terms of speech perception, there are two basic categories of experiments: *discrimination* and *identification*. The seminal research that gave rise to this dichotomy was conducted by Liberman, Harris, Hoffman & Griffith (1957), who hypothesized that the discrimination of certain speech sounds would be limited by their classification. Specifically, the ability of listeners to discriminate between two stimuli would be determined by the extent to which the listeners classified them differently. This led to the claim that speech sounds were perceived categorically: sounds which fell within a certain range on a continuum would be classified as belonging to the same category and a clear line could be drawn between sound categories. This meant that within-category differences were not perceivable (all sounds belonging to the same category would be pulled together under that label). Later research showed that categorical perception was not the whole story and that indeed listeners could distinguish sounds from within the same category. This in turn meant that listeners were able to distinguish sounds based upon their acoustic properties as well as their phonemic labels.

These different types of perception are often referred to as phonological or phonetic/acoustic perception, or language-specific vs. psycho-acoustic auditory perception (Babel & Johnson 2010). Language-specific perception is determined by phonological categories, that is, the contrastive sounds that are part of a language's inventory while auditory or phonetic discrimination, on the other hand, taps into the listener's ability to detect raw acoustic differences between two stimuli.

A discrimination experiment measures subjects' ability to differentiate stimuli.[2] Its most basic form is the AX or same-different task. The AX design involves the presentation of two stimuli per trial and the subject has to determine whether they are the same or different. In experiments of this type, the decision is made based upon the second member of the trial. Nonetheless, as McGuire (2010) points out, there are several problems with this task. For example, a bias emerges towards similar responses when the stimuli are very close perceptually. Another problem arises when the perceptual distance between the stimuli is different and the bias towards same responses is inconsistent.

The AX task data analysis usually involves either accuracy scores derived from percent correct/proportion correct (which require arcsine transform to account for potentially non-normal distribution) or a sensitivity measure such as *d'*. This

measure takes into account subject bias in responses accounting for accuracy (correct responses, or 'hits') and false positives ('false alarms') in one measure. In simple experimental designs, the *d'* calculation is based upon the z-transformation of the proportions (see Macmillan & Creelman 2005). A variant of the AX task involves speeded responses. The design is the same, but the ISI is very short – usually less than 500ms – with the objective of forcing listeners to rely on perception in the auditory mode rather than the phonetic (i.e. more abstract) mode. The advantage of a speeded AX task is that it allows researchers to tap into psychoacoustic differences and potentially by-pass speech categories. However, given the strict time constraint, some listener groups may have difficulty completing the task and this may lead to a high number of unusable responses. To combat this, McGuire (2010) recommends giving subjects immediate feedback on both accuracy and time, so as to provide a target for both.

A third methodological variant is the ABX task, in which the listener compares the A and B stimulus to the third and determines which is most similar. Problems with this design arise due to recency effects and possible bias towards B responses in addition to the added memory load because the target stimulus is no longer equidistant from the comparison stimuli. In an effort to address this, researchers have also used the AXB task, but this led to biases towards A responses (van Hessen & Schouten 1999). A variant on the ABX task is the two alternative forced choice task (2AFC) in which listeners are required to determine which stimulus came first, whether A or B. However, this task requires explicit labels for the categories tested and the listener must be able to perceive differences between them.

The second category of speech perception tasks involves identification or categorization experiments. Subjects hear a single stimulus (either from a closed set or open set, as in 'write what you hear,' McGuire 2010) and must apply a label to the stimulus. This type of task is used to tap into language-specific categories and is often applied to stimuli drawn from a continuum, where subjects have to determine where the category boundary lies. Related to categorization experiments are those where the listener is asked to judge the 'goodness' of a particular stimuli in terms of a particular target or in terms of stimuli previously played. This type of experiment is common for testing the hypotheses of the Perceptual Assimilation Model (Best 1995). Other tasks require subjects to make subjective evaluations in terms of the 'nativeness' of a certain stimulus or pair of stimuli using a scale. It is common practice to transform the data into z-scores to account for possible variability across subjects with respect to the scalar values.

3.2.1 Vowels

Spanish has a five-vowel system /i e a o u/ with no vowel reduction or laxing. For adult L1 Spanish/L2 English learners, perceiving the English tense-lax vowel contrast /i - ɪ/ as in 'heat' and 'hit' is notoriously difficult. In English, listeners

distinguish tense from lax vowels primarily based upon two dimensions: spectrum (vowel quality) and duration (vowel length), but the principle cue lies in spectral properties; vowel duration is secondary (Hillenbrand, Clark & Houde 2000). A large body of research has shown that Spanish listeners tend to rely upon vowel duration when identifying the English tense and lax vowels (Bohn 1995, Escudero 2006, Escudero & Boersma 2004, Kondaurova & Francis 2008), at least at the earliest stages. Escudero & Boersma (2004) examined how L1 Spanish listeners with varying levels of English experience shift their reliance upon one cue over the other and gradually move away from a total reliance upon duration and towards reliance on spectral changes. In terms of methodology, Escudero & Boersma created a two-dimensional frequency (F1 – duration) stimuli set and had native Spanish listeners carry out a categorization task on the individual vowel sounds as either 'sheep' or 'ship,' based upon pictures that appeared on the computer screen. Listeners were given as much time as necessary to select their response.

According to Escudero & Boersma, because duration is not a cue to vowel categories in their L1, Spanish listeners default to it when hearing English. In order to acquire the new categories, L1 Spanish listeners rely upon a statistical learning mechanism to track the distribution of duration in the input (see Boersma & Hayes 2001, Gradual Learning Algorithm for an algorithm of how such an input-driving constraint re-ranking might occur).[3] In other words, the acquisition of a previously non-discriminable contrast requires a re-weighting of acoustic cues that allows listeners to create a new category for their target language.

3.3 Speech production

In terms of possible methodology for production studies, researchers commonly have participants read lists of words, sentences or paragraphs, which afford the obvious advantage of tight control over phonetic context, speech rate and lexical items. There are obvious drawbacks, however, such as the artificiality of the task and the formality. The formal nature of reading a list out loud may lead to fewer examples of allophonic or phonetic variability than would be obtained in a more natural speech sample (Major 2001). Imitation of a native-speaker speech sample has also been used in various studies (Flege, Munro & MacKay 1995, Flege, Yeni-Komshian & Liu 1999), a method that may not get at learners' productive capabilities in a meaningful fashion. More authentic tasks, such as picture description or recounting a personal anecdote (Moyer 2007) while having the obvious advantage of authentic and ecologically valid methodology, draw upon more general language abilities which may impede a true evaluation of speaker proficiency. As pointed out by Moyer (2007) it is recommended that studies incorporate a variety of tasks in order to obtain reliable data when the objective is global pronunciation ratings. However, if the objective is to evaluate the production of a specific phonetic target, highly controlled experimental data collection methods

may be the only option. Thus, it is important to carefully consider the goals of the experiment itself.

3.3.1 Rhotics

Face (2006) examined the acquisition of the Spanish rhotics (*vibrante multiple* /r/ and *vibrante simple* /ɾ/ by native speakers of English. While English has only one rhotic, which is produced with either a bunched or retroflex articulation, Spanish has two, which are contrastively distributed in word medial position (orthographically distinguished as *rr* vs. *r*) but non-contrastive in word-initial position, where only the trill can occur. As predicted by Best & Tyler (2007), native English speakers do not have great difficulties with perceiving the two rhotics. Instead, the problem lies with articulation of the trill, in particular. Researchers have argued that the trill is the more 'marked' segment, as evidenced by the earlier acquisition of the tap and the trill's greater complexity – in terms of typological rarity and articulatory complexity.

Face (2006) examined the acquisition of intervocalic [r] by 41 L2 Spanish learners (fourth semester university students and Spanish majors or minors in an upper-level course) and a control group of native speakers. In terms of methodology, this study is particularly interesting because of the detailed consideration given to the different types of non-target segments produced by the L2 speakers. Face provides an analysis of the alternative productions L2 Spanish speakers use when not articulating the target trill and shows conclusively that these change over the course of greater experience with Spanish. For the two learner groups, the tap was produced more accurately than the trill and there were significant differences between both groups and the native speakers. The more advanced learner group overwhelmingly substituted the tap for the trill while the lower proficiency group substituted the English alveolar approximant.

This type of data and analysis is useful for research into precisely how learners change over the course of developing L2 articulatory routines and reminds investigators that binary 'correct' vs. 'incorrect' analyses are no longer sufficient to adequately capture changes in learners' L2 abilities. Another important finding is that native speakers did not perform uniformly in terms of their production of [r]: the native Spanish speakers reached 92% accuracy on the tap and only 86% accuracy on the trill, according to Face's phonetic characterization of target sounds. This suggests a sort of 'moving target' effect for L2 acquisition and questions the goal of a unitary phonetic model for L2 speakers. As stated above, the tasks used to elicit data play a very important role in precisely what type of data is ultimately obtained. In Face (2006), participants read a short story, which the author claims led them to use a less guarded, formal style (Face 2006, p.4). In Rose (2010), on the other hand, participants produced oral narratives elicited by means of two picture-

story description tasks. When interpreting the results from such studies, it is important to keep in mind the elicitation method used by the authors.

3.3.2 Stop consonants

Stop consonants, such as voiceless /p t k/ and voiced /b d g/, are among the most widely examined classes of sounds in L2 speech production studies. One of their primary articulatory characteristics is voice onset time, or VOT. VOT refers to the time between the release of the airflow (release burst) and the beginning of vocal cord vibration, or voicing. At the phonetic level, stops are classified into three categories, according to the average VOT duration: long-lag stops are produced with long VOT durations (generally longer than 35 milliseconds [ms.]), short-lag stops have short VOT durations (0-35 ms.), and prevoiced stops exhibit voicing throughout the closure and are therefore often expressed as having negative VOTs (Lisker & Abramson 1964). What is important for L2 acquisition purposes, however, is that while the same phonemic distinction may exist across languages (e.g. voiceless /p t k/ vs. voiced /b d g/), it differs with respect to the phonetic realization of those phonemes. In English, [p t k] are long-lag stops, while voiced [b d g] are short-lag stops. In Spanish, voiceless [p t k] are short-lag stops, while voiced [b d g] exhibit prevoicing. In a comparative study of both early and late learners, Flege (1991) examined the production of the L2 English voiceless stops /p t k/ by L1 speakers of Spanish. He found that early learners produced L2 /p t k/ with mean VOT durations that were similar to those for monolingual English speakers. Late learners, however, produced L2 /p t k/ with compromise VOT values. Flege concluded that the early learners had been able to establish separate phonetic categories for L1 Spanish and L2 English stops, while the late learners were inhibited by equivalence classification.

In the case of Spanish voiced stops, there is an additional issue that makes their acquisition by native English speakers potentially difficult. When they occur in intervocalic position, whether phrase or word-internally, or after [l], Spanish voiced stops are realized as approximants [β̞ ð̞ ɣ̞] (see work by (e.g. Hualde, Simonet, Shosted & Nadeu 2010, Martínez-Celdrán 2004, Romero 1995). This allophonic split causes learning difficulties for native English speakers primarily because a) they are allophones, which by definition do not change the meaning of the word and b) the two allophones are represented by the same orthographic symbols, which correspond to stops in English. Thus, this learning task is another example of how important it is for researchers to keep in mind the potentially different levels of representation involved in L2 speech acquisition and how orthography may modulate these effects.

There has been considerable research on the acquisition of these alternations by native English speakers of varying Spanish proficiency levels and under various task demands (e.g. Zampini 1994). Given that the approximant has greater intensity

than the plosive, when examining the production of the alternation, researchers typically compare the intensity (as root mean square values, see Lavoie 2001, for a more complete description) of the segments and calculate a ratio value to compare the plosive to the approximant. Hualde et al. (2010) also used spectral tilt and maximum velocity to examine the difference between the lenited segments and the plosives.

In a series of recent studies, Shea & Curtin (2010, 2011) show that knowledge of the context for the alternation plays a crucial role in whether or not learners produce the correct alternant. In other words, being able to produce and perceive allophones in a target-like fashion necessarily involves knowledge of the context that conditions each alternant. In methodological terms, this speaks directly to a need for testing the types of representations created during second language phonological/phonetic acquisition. One way researchers might do so is by means of cue-weighting studies (see below) or priming studies that can tap into lexical and acoustic representations (there is an extensive body of work on priming in Spanish-English bilinguals and L2 learners, see research by Kroll & colleagues, among others, on the bilingual lexicon).

It is difficult to completely separate segmental acquisition from the prosodic structure in which it is embedded, either at the foot/phrasal or utterance level. Indeed, prosodic structure drives phonological alternations in many languages, Spanish among them (see work by Cho and colleagues). Nonetheless, current research on second language segmental acquisition does not always take into account prosodic structure as a driving force behind phonological alternations and phonetic effects. [4]

In the following section I discuss the L2 acquisition of phonotactics, stress and intonation, or what are commonly called 'suprasegmentals.'

4. Suprasegmentals

4.1 Syllable structure

The role of phonotactics in second language speech has been widely studied (e.g. Davidson 2006) and it has been shown that second language learners repair consonant clusters that do not respect the phonotactics of their native language. However, the second language acquisition literature also provides ample evidence suggesting that speakers do not produce all unattested sequences with equal accuracy and native language phonotactics and universal sonority restrictions play a strong role in terms of the form repairs may take (Broselow & Finer 1991, Davidson 2006 and Major 2001).

Carlisle (1998, 1999) has carried out extensive research on the production of two and three-member onset clusters in English by native Spanish speakers (three-member clusters are illegal in Spanish). His work has shown that phonotactic

repairs by second language learners follow universal markedness principles, particularly the Sonority Sequencing Principle. In Davidson (2010), native English and Catalan speakers produced onset clusters that were unattested in their respective native languages. Clusters varied according to manner, voicing and place as well as the input modality – whether audio only or audio+text condition. In general, her results show a complex interaction between language universals and language-specific factors in terms of the type of phonotactic repair employed by second language learners and very little difference for the factor of input modes. The methodology employed by Davidson (repetition or reading) distinguishes this study from many others that rely purely upon either reading (Carlisle) or spontaneous speech. Indeed, the results from Davidson (2010) suggest that phonotactic repairs are consistent across task demands, which contradicts findings regarding individual segments. Further study is required to determine whether running speech may have an effect upon this. For example, re-syllabification effects in Spanish were found to effect e-epenthesis by native Spanish speakers producing s+C clusters in English by Carlisle (1999).

4.2 Stress

Guion, Harada & Clark (2004, replication and extension of Guion, Clark, Harada & Wayland 2004) examined the perception and production of stress in English by late and early Spanish-English bilinguals and how three factors that are purported to affect the placement of stress in Spanish and English influenced their performance: syllable weight, lexical class and stress patterns of phonologically similar words. Guion, Harada & Clark investigated how these three factors related to the placement of stress in English interacted with age of acquisition in the case of early and late L1 Spanish-L2 English bilinguals. The results of a logistic regression analysis showed that all three predictor variables (lexical class, syllabic structure and phonological similarity) made significant predictions of stress placement for the native English and early Spanish-English bilinguals. For the late bilinguals, only lexical class and phonological similarity were significant, suggesting that late learners are not sensitive to phonological structure when assigning stress. These results indicate that both early and late bilinguals use distributional information found in the input to determine a probabilistic placement of stress on English non-words. Moreover, the fact that differences were observed between the early and late groups in terms of syllable weight suggests that this type of rule generalization may be more difficult to learn.

In methodological terms, Guion, Harada & Clark's study is particularly interesting because they integrate perception and production data in a study that tests three different models of stress placement (see Cabrelli-Amaro, this issue, for more details on the methodology). Moreover, by including early and late bilinguals the authors were able to test whether both groups use the same type of information

in the placement of stress. As their results suggest, it appears that early bilinguals can make use of the abstract phonological knowledge which the late learners do not. This sheds light on the Critical Period debate and the nature of post-CP knowledge. In methodological terms, research that teases apart how learners of different proficiency levels use different information in the input contributes to how the CP in language learning should be conceptualized (see Shea & Curtin 2010 for a similar argument).

4.3 Intonation

In a well-controlled study on the development of L2 Spanish intonation in a study-abroad context, Henrickson, Geeslin & Willis (2010) used a computerized production task to elicit declaratives, absolute interrogatives and pronominal interrogatives from five English-speaking learners studying abroad in León, Spain. The authors presented participants (n = 5) with 162 PowerPoint slides which contained stories that served to contextualize the lexical items and sentences used in the experimental task. The second part of the study involved reading a series of target sentences that were embedded in a context that served to elicit the particular structure, whether declarative, absolute interrogative or pronominal interrogative. Data was collected at two different points during the study abroad session. Henrickson, Geeslin & Willis' findings revealed two patterns in terms of learner development (p.39): increased use of a given set of intonation strategies over time, or increased use of intonation strategies (i.e. pitch accents and boundary tones) over time.

5. Training studies

Another area of research in L2 phonological/phonetic acquisition involves training listeners (more than speakers) to perceive previously assimilated L2 sounds. In most training studies, participants are given a pre-test to obtain a baseline perceptual measurement and then a post-training test to determine the effects of the training regimen. Training sessions can be numerous, depending upon the goal of the study and often require various sessions to complete. In general, phonetic training has focused on how to make listeners create a new category for a previously assimilated L2 sound. Two processes have generally been used: cue enhancement (non-native listeners learn to perceive cues which are relevant to native speakers in detecting a contrast) and cue inhibition (non-native listeners learn to ignore cues that are not relevant to native speakers). Essentially, both approaches are based in attention-to-dimension models (Francis & Nusbaum 2002): when listener attention to a particular dimension is enhanced, tokens separated by that dimension become more discriminable. When listener attention is reduced, tokens become less discriminable. Weighting within and between perceptual dimensions optimizes performance in the

native language but can interfere with the establishment of the weightings necessary for a second language (Guion & Pederson 2007).

In a study directly related to Spanish, Kondaurova & Francis (2010) tested three different training methods in order to determine which is most appropriate for training L1 Spanish listeners to perceive the English /i/-/ɪ/ contrast. After a pretest to select participants who relied upon the duration cue primarily, subjects participated in one of three training regimes, based upon enhancement, inhibition or natural cue distribution. Results show that while all three training groups improved their perception of the contrast according to a baseline measurement, the inhibitory training paradigm was more effective than the other two and the least effective was the natural distribution training. These results from Kondaurova & Francis suggest a) that second language learners can be trained to pay more attention to category dimensions that are not relevant for discrimination in their native language and b) training on ignoring dimensions is more beneficial than training on enhancing previously ignored dimensions.

6. Conclusions

The field of second language phonetics and phonology enjoys a large and well-established research tradition. As can be seen from the variety of studies and methodologies discussed in this article (and others, such as neurolinguistic and articulatory methods which were not addressed) researchers have explored adult second language acquisition from a wide range of perspectives. Going forward, there is a need for more work focused on incorporating the data into a broader model of second language acquisition as a process (i.e. mechanisms involved) and the nature of the representations created based upon interactions with the target language. As well, I join Cabrelli-Amaro (this volume) in calling for studies on the acquisition of a third language. The field will also benefit from more well-controlled research on the relationship between the L1 and an L2 acquired in childhood, as in the case of Spanish heritage speakers. This population provides an interesting source of data on how first language phonology and phonetics change and shift when in contact with a second language. In conclusion, the field of L2 phonology is growing rapidly, assisted by the widespread availability of free software programs such as PRAAT that allow investigators to ask (and answer) ever more sophisticated research questions.

Acknowledgments

Thank you to Tim Gupton for organizing these contributions and for editing advice.

Notes

1 I will not address bilingual acquisition, L1 acquisition, child L2 acquisition or heritage speaker phonological/phonetic knowledge, although these are very interesting and growing fields of inquiry.
2 This section is based on McGuire (2010), which provides a more detailed elaboration and theoretical justification of experimental methodology commonly used in speech perception research.
3 Morrison (2008, 2009) contends that because Spanish listeners' responses correlate positively with duration and negatively with spectral properties, there is in fact a stage of perception that precedes the duration-based stage suggested by Escudero & Boersma.
4 There is a large body of research examining domain-initial/final effects that incorporates stress and prosodic structure into the analysis of segmental production data. Examining the role of this in L2 acquisition is starting to gain ground, particularly in the area of speech segmentation and lexical recognition (see work by Cho 2011).

References

Abrahamsson, Niclas & Kenneth Hyltenstam. 2009. Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. Language Learning 59, 249-306.

Archibald, John. 1993. The learnability of English metrical parameters by adult Spanish speakers. International Review of Applied Linguistics 31, 129-141.

Archibald, John.1994. A formal model of learning L2 prosodic phonology. Second Language Research 10, 215-240.

Babel, Molly & Keith Johnson. 2010. Accessing psycho-acoustic perception with speech sounds. Laboratory Phonology 1, 179-205.

Best,Catherine. 1995. A direct realist view of cross-language speech. In Winnifred.Strange (ed.), Speech perception and linguistic experience, 171-204. Baltimore: York Press

Best, Catherine & Michael Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn & Murray Munro (eds.), Language experience in second language speech learning: In honor of James Emil Flege, 13-34. Amsterdam: John Benjamins.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. Linguistic Inquiry 32, 45-86.

Bohn, Ocke-Schwen. 1995. Cross-language speech perception in adults. In Winifred Strange (ed.), Speech perception and linguistic experience, 275-300. Timonium, MD: York Press.

Bohn, Oke-Schwen & Murray Munro (eds). 2007. Language Experience in second language speech learning: In honor of James Emil Flege. New York: John Benjamins Publishing Company.

Bongaerts, Theo. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In David Birdsong (ed.), Second language acquisition and the critical period hypothesis, 133-159. Mahwah, NJ: Lawrence Erlbaum Associates.

Boomershine, Amanda, Kathleen Currie Hall, Beth Hume & Keith Johnson. 2008. The influence of allophony vs. contrast on perception: The case of Spanish and English. In Peter Avery, B. Elan Dresher & Keren Rice (eds.), Phonological contrast, 143-172. Berlin: Mouton.

Broselow, Ellen & Daniel Finer. 1991. Parameter setting in second language phonology and syntax. Second Language Research 7, 35-59.

Bundgaard-Nielsen, Rikke, Catherine Best & Michael Tyler. 2011. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. Journal of Applied Psycholinguistics 32, 51-67.

Carlisle, Robert S. 1998. The acquisition of onsets in a markedness relationship: A longitudinal study. Studies in Second Language Acquisition 20, 245-260.

Carlisle, Robert S. 1999. The modification of onsets in a markedness relationship: Testing the Interlanguage Structural Conformity Hypothesis. Language Learning 49, 59-93.

Cho, Taehong. 2011. Laboratory phonology. In Nancy C. Kula, Bert Botma & Kuniya Nasukawa (eds.), The continuum companion to phonology, 343-368. London: Continuum.

Davidson, Lisa. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. Journal of Phonetics 34, 104-137.

Davidson, Lisa. 2010. Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. Journal of Phonetics 38, 272-288.

DeKeyser, Robert M. 2000. The robustness of critical period effects in second language acquisition. Studies in Second Language Acquisition 22, 499-533.

DeKeyser, Robert M., Iris Alfi-Shabtay & Dorit Ravid. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. Applied Psycholinguistics 31, 413-438.

Escudero, Paola & Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. Studies in Second Language Acquisition 26, 551-585.

Face, Timothy L. 2006. Intervocalic rhotic pronunciation by adult learners of Spanish as a second language. In Carol A. Klee & Timothy L. Face (eds.), 7th Conference of the acquisition of Spanish and Portuguese as first and second Languages, 47-58. Somerville, MA: Cascadilla Proceedings Project.

Flege, James E. 1991. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. Journal of the Acoustical Society of America 89, 395-411.

Flege, James E. 1995. Second-language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), Speech perception and linguistic experience: Issues in cross-language research, 229-273. Timonium, MD: York Press.

Flege, James E. 2002. Interactions between the native and second-language phonetic systems. In Petra Burmeister, Thorsten Piske & Andreas Rohde (eds.), An integrated view of language development: Papers in honor of Henning Wode, 217-244. Trier: Wissenschaftlicher Verlag.

Flege, James E. 2007. Language contact in bilingualism: Phonetic system interactions. In Jennifer Cole & José I. Hualde (eds.), Laboratory Phonology 9, 353-380. Berlin: Mouton.

Flege, James E., Ocke-Schwen Bohn & S. Jang, S. 1997. Effects of experience on non-native speakers' production of English vowels. Journal of Phonetics 25, 437-470.

Flege, James E. & Serena Liu. 2001. The effect of experience on adults' acquisition of a second language. Studies in Second Language Acquisition 23, 527-552.

Flege, James E., Murray J. Munro & Ian R.A. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. JASA 97, 3125-3134.

Flege, James E., Grace Yeni-Komshian & Serena Liu. 1999. Age constraints on second language learning. Journal of Memory and Language 41, 78-104.

Francis, Alexander & Howard C. Nusbaum. 2002. Selective attention and the acquisition of new phonetic categories. Journal of Experimental Psychology: Human Perception and Performance 28, 349-366.

Guion, Susan, J. J. Clark, Tetsuo Harada, & Ratree P. Wayland. 2004. Early and late Spanish–English bilinguals' acquisition of English word stress patterns. Bilingualism: Language and Cognition 7, 207-226.

Guion, Susan, Tetsuo Harada & J. J. Clark. 2004. Early and late Spanish–English bilinguals' acquisition of English word stress patterns. Bilingualism: Language and Cognition 7, 207-226.

Guion, Susan & Eric Pederson. 2007. Investigating the role of attention in phonetic learning. In Ocke-Schwen Bohn & Murray J. Munro (eds.), Second language speech learning: The role of language experience in speech perception and production: A festschrift in honor of James E. Flege, 57-77. Amsterdam: John Benjamins.

Henriksen, Nicholas C., Kimberly L. Geeslin & Erik W. Willis. 2010. The development of L2 Spanish intonation during a study abroad immersion program in León, Spain: Global contours and final boundary movements. Studies in Hispanic and Lusophone Linguistics 3, 113-162.

Hernandez, Arturo E. & Iliana Reyes. 2002. Within- and between-language priming differ: Evidence from repetition of pictures in Spanish-English bilinguals. Journal of Experimental Psychology: Learning, Memory and Cognition 28, 726-734.

van Hessen, Arjan J. & Bert M.E.H. Schouten. 1999. Categorical perception as a function of stimulus quality. Phonetica 56, 56-72.

Hillenbrand, James, Michael J. Clark & Robert A. Houde. 2000. Some effects of duration on vowel recognition. Journal of the Acoustical Society of America 108, 3013-3022.

Hualde, José I., Miquel Simonet, Ryan Shosted & Marianna Nadeu. 2010. Quantifying Iberian spirantization: Acoustics and articulation. Presented at LSRL 40, Seattle, WA.

Keating, Patricia A., Taehong Cho, Cecile Fougeron & Chai-Shune Hsu. 2003. Domain-initial strengthening in four languages. In John Local, Richard Ogden & Rosalind Temple (eds.), Papers in laboratory phonology 6: Phonetic interpretations, 145-163. Cambridge, UK: Cambridge University Press.

Kondaurova, Maria V. & Alexander L. Francis. 2010. The role of selective attention in the acquisition of English tense and lax vowels by native Spanish speakers: Comparison of three training methods. Journal of Phonetics 38, 569-587.

Lavoie, Lisa. 2001. Consonant strength: Phonological patterns and phonetic manifestations. New York: Garland Publishing Inc.

Lenneberg, Eric H., 1967. Biological foundations of language. New York: Wiley.

Liberman, Alvin M., Katherine S. Harris, Howard S. Hoffman & Belver C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54, 358-368.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-linguistic study of voicing in initial stops: Acoustical measurements. Word 20, 384-422.

Macmillan, Neil A. & C. Douglas Creelman. 2005. Detection Theory: A User's guide (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Major, Roy. C. 2001. Foreign accent: The ontogeny and phylogeny of second language phonology. New Jersey: Lawrence Erlbaum Associates.

Martínez-Celdrán, Eugenio. 2004. Problems in the classification of approximants. Journal of the International Phonetic Association 34, 201-210.

McGuire, Grant. 2010. A brief primer on experimental designs for speech perception research. In Grant McGuire (ed.), The UCSC Linguistics Research Center 2010 laboratory report, 93-120. Santa Cruz, CA: LRC Publications.

Morrison, Geoffrey S. 2008. L1-Spanish speakers' acquisition of the English /i/–/I/ contrast: Duration-based perception is not the initial developmental stage. Language & Speech 51, 285-315.

Morrison, Geoffrey S. 2009. L1-Spanish speakers' acquisition of the English /i/–/I/ contrast II: Perception of vowel inherent spectral change. Language & Speech 52, 437-462.

Moyer, Alene. 1999. Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. Studies in Second Language Acquisition 21, 81-108.

Moyer, Alene. 2004. Age, accent and experience in second language acquisition: An integrated approach to critical period inquiry. Clevedon: Multilingual Matters.

Moyer, Alene. 2007. Empirical considerations on the age factor in L2 phonology. Issus in Applied Linguistics 15, 109-127.

Pierrehumbert, Janet. 2003. Phonetic diversity, statistical learning, and acquisition of phonology, Language and Speech 46, 115-154.

Romero, Joaquin. 1995. Gestural coordination in Spanish: An experimental study of spirantization and aspiration. Doctoral dissertation, The University of Connecticut.

Rose, Marda. 2010. Intervocalic tap and trill production in the acquisition of Spanish as a second language. Studies in Hispanic and Lusophone Linguistics 3, 379-419.

Shea, Christine & Suzanne Curtin. 2010. Discovering the relationship between context and allophones in a second language: Evidence for distribution-based learning. Studies in Second Language Acquisition 32, 581-606.

Shea, Christine & Suzanne Curtin. 2011. Experience, representations and the production of second language allophones. Second Language Research 27, 229-250.

Simonet, Miquel. 2011. Technology in phonetic science: Setting up a basic phonetics laboratory. Studies in Hispanic and Lusophone Linguistics 4, 557-576.

Simonet, Miquel. 2012. The L2 acquisition of Spanish phonetics and phonology. In José I. Hualde, Antxon Olarrea & Erin O'Rourke (eds.), Handbook of Hispanic linguistics. London: Wiley-Backwell.

Tremblay, Annie. 2009. Phonetic variability and the variable perception of L2 word stress by French Canadian listeners. International Journal of Bilingualism 13, 35-62.

Zampini, Mary L. 1994. The role of native language transfer and task formality in the acquisition of Spanish spirantization. Hispania 77, 470-481.