

# Empirical Considerations on the Age Factor in L2 Phonology

Alene Moyer

University of Maryland

*Phonological skill is widely regarded as subject to a critical period for language learning, though the nature of relevant maturational changes has yet to be clarified. Theoretical and empirical research on phonological skill development among late learners has confirmed several observable influences on short- and long-term attainment; however, these research traditions have rarely acknowledged contextual influences, much less the learner's role in the process. In this paper, I outline a number of methodological concerns for current research, and provide specific recommendations regarding participant selection, tasks, ratings and raters, factors tested, and analyses, in order to better account for influences on phonological attainment that co-vary with age.*

## INTRODUCTION: AGE OF ONSET AND ULTIMATE ATTAINMENT IN L2 PHONOLOGY

The question of a critical period for language learning has enjoyed renewed interest of late in the face of evidence that shows inconsistent age effects across tasks that measure grammar skills and pronunciation in a second language (L2) (Birdsong, 1992; Bongaerts, Summeren, Planken, & Schils, 1997; DeKeyser, 2000; Flege, Yeni-Komshian, & Liu, 1999; Ioup, Boustagi, El Tigi, & Moselle, 1994; White & Genesee, 1996 – see DeKeyser & Hall, 2005 and Hyltenstam & Abrahamsson, 2003, for reviews; see Long, 2005, for critique of counter-evidence). In the realm of phonology specifically, the evidence overwhelmingly confirms age-related declines in the ability to perceive and produce new sounds (Bohn & Flege, 1992; Flege, 1991; Flege & Hillenbrand, 1987; Flege & MacKay, 2004; Moyer, 1999, 2004, 2007a; Pallier, Bosch, & Sebastian-Galles, 1997; cf. Bongaerts, Planken, & Schils, 1995; Ioup et al., 1994), which is in keeping with Lenneberg's (1967) original assertion that of all language skills, phonology would be the most difficult to attain after puberty (i.e., following the close of any putative critical period for language learning).

Considering how little is known regarding the specific neurological and cognitive changes that presumably affect the perception and production of new (L2) sounds in negative ways, much more can and should be done to explore how individual constraints operate, e.g., aptitude for language learning and/or oral mimicry; desire to sound native; access to instruction and feedback on pronunciation, and so on. Nearly 20 years ago, Tarone (1987) wrote that “complex interrelationships of language, mind, body and society” operate in tandem to determine phonological attainment in L2 (p. 84). With this in mind, some researchers have

begun to explore individual factors among late learners, such as formal instruction, motivation, and opportunities for authentic L2 contact as they impact phonological authenticity (Bongaerts et al., 1997; Diaz-Campos, 2004; Elliott, 1997; Flege et al., 1999; Moyer, 1999, 2004, 2006, 2007a, 2007b; Piller, 2002), with the evidence pointing to the significance of these factors as well, though none appears to be as impressive, statistically speaking, as the age of onset (AO) factor itself. Much remains to be done here, nevertheless, since so many such factors co-vary with age and do figure impressively in recent statistical models (see Moyer, forthcoming, for discussion).

Assuming that phonology, like any other skill in L2, is subject to both universal mechanisms and individual influences, it is appropriate to reflect on whether our methods have adequately measured this complex convergence of factors up to now. This paper examines the breadth and validity of the instrumentation in age effects studies in L2 phonology in order to note where we can strengthen our approaches (see also Long, 2005). It could be argued that many studies in second language acquisition (SLA) cannot be expected to adhere to strict standards of validity<sup>1</sup> because our populations are often specialized and our questions targeted to unique circumstances. This discussion therefore intends only to raise questions and suggest guidelines for future work that could enhance validity, and thereby, the reliability of findings. To this end, I introduce several methodological concerns regarding participant selection, task design, ratings and rater reliability, and data analyses. Throughout, I emphasize the need for, and viability of, more carefully designed instruments to address the impact of age vis-à-vis other factors. Following this introduction of the problems, I outline several methodological recommendations for future investigations.

## METHODOLOGICAL ISSUES

### **Participant Selection and the Ultimate Attainment Issue**

In studies on perception and production of new L2 sounds, participants may be new, inexperienced learners, or they may be experienced learners who still have difficulties with pronunciation (Derwing & Rossiter, 2003; Dupoux, Peperkamp & Sebastián-Gallés, 2001; Olson & Samuels, 1982; Snow & Hoefnagel-Hoehle, 1982; Strange & Dittmann, 1984). More commonly, studies of age effects look at the long-term impact of neurobiological age on language learning.<sup>2</sup> In other words, the focus is not so much on the process as the end-state, or “ultimate,” attainment point. With this in mind, the sample is often chosen from among immigrant populations residing in-country for several or many years (Flege, Munro, & MacKay, 1995; Yeni-Komshian, Flege, & Liu, 2000; Yeni-Komshian, Robbins, & Flege, 2001). On the one hand, long-term learners provide an opportunity to test age effects against other factors; on the other, their experience reveals widely varying opportunities for authentic interaction and practice, widely varying orientations to the target language, and widely varying opportunities to receive L2 instruction. In fact, those

with extensive L2 experience are sometimes recruited for their exceptional degree of motivation towards the target language (TL) (Bongaerts, Planken, & Schils, 1995; Moyer, 1999, 2004), or are pre-selected by teachers or linguists specifically for their excellent pronunciation (Bongaerts et al., 1995, 1997).

Viewed across, and even within, studies, participant backgrounds are seldom parallel. These inconsistencies are problematic on several grounds. Most obviously, there is the problem of comparability (and therefore generalizability) of results. Concerns about validity and reliability<sup>3</sup> are not far behind. As Porte (2002) points out, selection bias can be an unwitting byproduct of studies that divide the sample into groups – such as in experimental research that involves a treatment. In studies of ultimate attainment, no treatment is involved, but we face a similar danger of selection bias. Subjects are often chosen for their specialized abilities, resulting in a homogeneous sample, which inherently affects generalizability since the mean and variance of their performance are likely to be distorted. Let us look at a specific example: In ultimate attainment studies, the goal is often to produce evidence counter to the (widely accepted) critical period for language learning. Is this best accomplished by selecting only exceptional cases? If so, should ‘exceptional’ status be determined by (a) self-report; (b) the researchers’ own assessment; or (c) the judgment of a teacher? All of these selection criteria have been used, though none is ideal. What are the odds that another researcher can replicate the same specialized population according to the same selection criteria?

Of further concern is the fact that selecting a very specialized group (e.g., advanced learners – often targeted for critical period studies because their exposure is extensive) will likely mean that some participants have a real advantage over others, often in terms of previous instruction, possibly even overt training on the perception and/or production of new (L2) sounds. The effects of widely varying instruction are most difficult to control for if we hope to establish equivalence of pedagogical focus, language content, classroom hours, etc. Unless such experiential variables are isolated specifically in the analysis (i.e., analyzed separately from the impact of AO), comparability within the sample of learners is in question (an example here is Birdsong’s 1992 study on grammaticality judgments where instructional experience is not adequately accounted for).

Across studies, comparability is clearly of great concern as well. For example, some studies incorporate an experimental training phase to address problematic L1-L2 contrasts for perception or production (or both) (Derwing & Rossiter, 2003; Dickerson, 1975, 1987; Dupoux et al., 2001; Olson & Samuels, 1982; Perani et al., 1998; Snow & Hoefnagel-Hoehle, 1982; Strange & Dittmann, 1984), then test participant performance following the treatment. Other studies test instruction as a long-term, general measure (Bongaerts et al., 1997; Moyer, 1999, 2004). The problem is that instruction can show differential benefits depending on how long it lasts, or the point in the process at which it is introduced. Derwing, Munro, and Wiebe (1997) find no lasting benefits for learners who had been immersed for many years (10 or more) and underwent a brief pronunciation course, but Bongaerts et al.

(1997), Elliott (1997) and Moyer (1999, 2004) find that overall duration of instruction is significant, with suprasegmental training specifically leading to more native-like ratings (Moyer, 1999 – see also Derwing & Rossiter, 2003; Missaglia, 1999 cited in Piske, MacKay, & Flege, 2001). Suter's 1976 study shows significance for long-term instruction, but when reanalyzed through factor analysis and regression (Purcell & Suter, 1980), this variable does not appear to be salient, especially when compared to living with native speakers and concern for pronunciation accuracy. The point is that if we do not account for the relevant experience of participants, we cannot confidently ascribe age effects to neurobiological age only. The intervention of something as (statistically) significant as overt instruction<sup>4</sup> must be ruled out – or controlled for – in accordance with the objectives of the study.

In addition to instruction, other measures of experience are significantly related to phonological performance, for example, length of residence (Asher & Garcia, 1969; Flege et al., 1995, 1999; Moyer, 2004, 2007a; Oyama, 1976; Thompson, 1991; see Piske et al., 2001, for discussion). One important consideration here is that length of residence (LOR) is a very rough measure of experience with little explanatory value on its own; there is no indication of learner engagement therein, such as the particulars of L2 usage or domains for contact (see Moyer, forthcoming), for example, frequency of contact, quality of input, etc. Experience has been measured as actual *use* of L2 relative to L1, first by Suter (1976) and Purcell and Suter (1980), and more recently by Flege et al. (1995, 1999) and Flege and MacKay (2004). The best indications are that minimal L2 use, or to put it more pointedly, *continued L1 use* appears to have a negative impact on L2 phonetic perception and production, regardless of how long one resides in the TL-speaking environment.

Attempting a comprehensive approach to the experience problem, Moyer (2004, 2007a, 2007b) has recently analyzed instruction, contact to native speakers, domains for L2 use, and actual (weekly) time spent using L2 relative to L1, in both quantitative and qualitative terms. Results point to a balance of socio-psychological and exposure-type variables for predicting accent, including contact with native speakers, length of residence and age of onset, as well as intention to reside in the TL-speaking environment permanently or long-term, comfort with assimilation to the TL culture, desire to improve accent, and sense of overall fluency. According to these studies, AO is not the most impressive factor statistically speaking, when compared in multiple regression analyses to other highly significant experiential factors (in Moyer's 2007a study, it accounted for far less of the variance than LOR). Given the recent confirmations of its significance, the experience construct – especially as a complex indicator of language use and orientation – deserves much more careful attention in our instruments.

Finally, factors of a socio-psychological nature should be taken into account for any learners beyond the critical period. This is especially important if we hope to view language acquisition more holistically (i.e., as an active process marked by conscious decision-making, not just subject to the impact of AO). There is mounting evidence that affect in general is closely related to cognition, affecting

short-term memory in language processing (see DeWaele, 2002; Pulvermüller & Schumann, 1994; Schumann, 2001). Recent research in L2 phonology confirms the significance of affective factors and learner orientation to the target language (see Dörnyei & Skehan, 2003, for discussion). Statistically significant relationships are evident between closer-to-native pronunciation abilities and motivation (Bongaerts et al., 1997; Flege et al., 1999; Moyer, 1999, 2004; cf. Oyama, 1976; Thompson, 1991), concern for pronunciation accuracy (Elliott, 1995a, 1995b; Purcell & Suter, 1980), attitudes toward the TL culture (Major, 1993; Stokes, 2001), satisfaction with attainment, and self-rating of accent (Moyer, 2004, 2006), and aptitude for oral mimicry (Purcell & Suter, 1980; Thompson, 1991). It is noteworthy that while these factors cannot be shown unequivocally to directly cause certain outcomes, neither can neurobiological age itself (DeKeyser & Hall, 2005). What we can see through qualitative analysis is that such factors lead to strategic behaviors that impact phonological fluency, for example, consciously seeking phonological feedback at the segmental and suprasegmental levels (Moyer, 2004).

### **Elicitation Techniques and Tasks**

Studies on accent in a second language differ in terms of elicitation techniques for production skills,<sup>5</sup> with some studies choosing only one of a number of possible techniques, and a few combining several. Following perhaps the most simplistic approach to the most basic question, *can any late learner perform at a native level under any circumstances?*, a number of studies rely on direct imitation or repetition after a native speaker (NS) (Flege et al., 1995, 1999; Markham, 1997; Olson & Samuels, 1982; Snow & Hoefnagel-Hoehle, 1982). Imitation as a means to discover whether a learner can ‘sound native,’ if only for a few seconds at a time, hardly provides insight into phonetic/phonemic accuracy under real-time processing conditions. More often, studies ask for autonomous production of isolated words, sentences, and/or paragraphs with no model or auditory input (Asher & Garcia, 1969; Bongaerts et al., 1995, 1997; Elliott, 1995a, 1995b; Moyer, 1999, 2004; Neufeld, 1980; Oyama, 1976; Thompson, 1991). Unrehearsed, spontaneous speech is arguably the most appropriate way to probe true ultimate attainment. It is typically guided by various stimuli, including picture description (Elliott, 1995a, 1995b; Ishida, 2003; Moyer, 2007a), and/or recounting personal experiences based on a prompt (Bongaerts et al., 1995, 1997; Fathman, 1975; Moyer, 1999, 2004, 2007a, 2007b; Suter, 1976). The drawback here is that these more authentic tasks inherently involve lexical, morphological, syntactic, and pragmatic abilities as well, and thus are ‘messier’ in terms of interpreting the ratings that result.

The validity concerns for these varied approaches to task design are obvious. Depending on the naturalness of a given task, a speaker/learner may be perceived as more or less authentic to a rater. Imitation tasks and reading aloud isolated words (or even phones/phonemes) may elicit closer-to-target production than free speaking, possibly due to a greater focus on accuracy in decontextualized tasks (Dickerson, 1975; Moyer, 1999; Tarone, 1982; cf. Moyer, 2004; Oyama, 1976; Sato, 1985).

But are these decontextualized tasks really an authentic representation of what the learner can do (i.e., in real communication)? In their defense, isolated words and sentences do have their place; word elicitation can check for accurate stress at the word-level, while sentence tasks can test phrasal features like intonation and rhythm. Free speech is arguably the most natural task, but it is also the most prone to mistakes beyond the phonological level.

To minimize some of these inherent limits and artificialities, some studies combine tasks, allowing for a comparison of pronunciation authenticity across task type. Early on, Dickerson (1975) and Tarone (1982) asserted that different tasks elicit varying degrees of authenticity based on the learner's relative (overt) focus on accuracy. In other words, where production is decontextualized – more controlled, such as in word lists – the participant may be better able to focus on phonetic precision. Whether controlled tasks really do lead to greater accuracy/authenticity compared to more open-ended ones is up for debate at this point since task comparisons have produced mixed results (see Bongaerts et al., 1995; Dickerson, 1987; Moyer, 1999, 2004, 2007a, 2007b; Oyama, 1976; Sato 1985; Thompson, 1991). More often than not, analyses of task effects are complicated by the fact that phonological skills are difficult to isolate from other language skills once the participant moves beyond the word-level. Moreover, we have little data on differential abilities within the phonological realm (e.g., segmental vs. suprasegmental) for the same study.

Additional task concerns include the number of tokens per task, the extent to which tokens challenge the participants' abilities, and the allowance for practice of actual tokens. To a large extent, the number of tokens is logically tied to the task itself. (No one expects a participant to provide 20 samples of free speech, but reading 25 words is certainly reasonable.) As of this writing, there is no clear formula for determining a minimum number of task tokens; anywhere from 10 to 100 words is possible in a word list. Raising serious internal and external validity concerns – as far as this researcher is concerned – some study designs purposefully avoid difficult features (Flege et al., 1995) though others purposefully include difficult tokens (based on L1-L2 contrasts) to verify whether advanced learners can perform as natives (Asher & Garcia, 1982; Bongaerts et al., 1997; Moyer, 1999, 2004; Thompson, 1991). Of even greater concern is the fact that some studies allow for practice, meaning that participants read the same token 2-3 times, and only the “best” or final instantiation is rated (Bongaerts et al., 1997; Flege et al., 1995; Yeni-Komshian et al., 2000). Others do not allow for any rehearsal in order to avoid practice effects (Moyer, 1999, 2004, 2007a, 2007b; see Long, 2005, for additional relevant points). Overall, more attention to internal validity is needed in task design if we hope to compare, and ultimately to generalize, findings.

### **Raters and Rating Criteria**

Studies on L2 accent typically rely on either spectrographic analysis to compare non-native performance to that of a native speaker baseline,<sup>6</sup> or ask native

speakers to judge non-native speech samples, usually for degree of accent. These native speakers are considered “experts,” even though their educational, social and regional background can vary widely within and across studies. Some are teachers, some are linguists, others have no formal training in linguistics or language education. Formal analyses of who is actually most sensitive to, or most critical of, non-native accent are contradictory.<sup>7</sup> For this reason, a balance of rater types may be best, since both experienced and inexperienced raters bring different levels of sensitivity to their respective judgments (Piske et al., 2001).

The reliance on native speakers has been criticized of late, as perpetuating the myths that: (a) all native speakers of a language sound alike and share the same intuitions about their L1, meaning that they must all be ‘pure’ monolinguals (Cook, 2002); (b) native speakers have the last word on who meets the standard ideal, which denies the reality that multiple standards can and do exist, for example, for world Englishes (see also Davies, 2003). These critiques notwithstanding, the reliance on native speakers as raters is convenient, if not logical, for the reason that they do appear to exercise a similar set of perceptions on what is *comprehensible*, if not accented, and these impressions are extremely useful if we assume that there is a ‘core’ of linguistic features that must be minimally accurate in order for communication to proceed (see Jenkins, 2002; Seidlhofer, 2004). The problem here is that the definition of native speaker – used as if it were universally understood – is usually limited to one regional (and ethnic) group, thereby excluding the practical reality that native speakers of a given language do not share an identical set of intuitions (see Sorace, 1996). Moreover, limited views on whose intuitions constitute the ideal inevitably exclude other L2 users (see Kachru & Nelson, 1996; Moyer, 2007b).

Perhaps even more pressing than the question of *who* is the question of *how many*. The number of raters across studies has ranged anywhere from 1 to 6, with no explanation for the number chosen. Since 2 to 3 raters are enough to get solid results from an inter-rater reliability measure, such as Cohen’s Kappa, this could be considered a good (minimum) standard (e.g., Moyer, 1999, 2004, 2007a, 2007b).

A rating scale is the most common system for judging native-ness, ranging from 4 (Asher & Garcia, 1969) to 9 points (Flege et al., 1999), with 5 to 6 being the most frequently employed range (e.g., 1= definitely native/no foreign accent ... 6= definitely non-native/strong accent) (Bongaerts et al., 1995, 1997; Elliott, 1995a, 1995b; Fathman, 1975; Moyer, 1999, 2004, 2007a, 2007b; Olson & Samuels, 1973; Oyama, 1976; Patkowski, 1980; Purcell & Suter, 1980; Snow & Hoefnagel-Hoehle, 1982; Thompson, 1991). Other options have been used as well: Markham (1997) and Neufeld (1980) have used categorical ratings; Moyer (1999, 2004) has used a binary judgment (native/non-native) combined with a confidence scale of 1-3 to end up with a 6-point scale; Major and Kim (1999) compared categorical with scalar ratings across two rater groups to look for discrepancies and overlaps in judgments. Whether the absence of an accepted, unitary scale compromises validity and reli-

ability of results has hardly been noted thus far (cf. Piske et al., 2001),<sup>8</sup> possibly because this aspect of instrumentation is relatively consistent compared to others treated here. Nevertheless, it is another aspect of the instrumentation that should be carefully considered because of its implications for comparability.

### **Factors and Analyses**

I have argued here (and elsewhere) for a more purposeful accounting for the interplay between AO and its co-varying influences in order to better understand age effects in L2 phonology. If there is one obvious weakness in the critical period research in SLA, it is a narrow approach to age effects, meaning that too few variables have been tested against age itself. This weakness has been recognized recently (Bialystok & Hakuta, 1999; Marinova-Todd, Marshall, & Snow, 2000; Moyer, forthcoming), and L2 phonological research is arguably doing a good deal to counter it.

To test for the strength of variables that co-vary with age (instruction, motivation, etc.), correlation analyses such as Pearson-Product Moment correlations are most common, followed by more sophisticated tests of multiple factors as they contribute differentially to predicting outcomes, (e.g., ANOVA and multiple regression analyses) (Flege et al., 1999; Moyer, 1999, 2004, 2007a, 2007b; Munro & Derwing, 1999, 2001; Oyama, 1976; Purcell & Suter, 1980; Thompson, 1991). The inclusion of such tests is significant, since only by measuring a range of potential influences, and applying complex tests like multiple regression, can the *relative* impact of various (age-related) factors be understood. Non-linguistic instruments must therefore be broad enough to measure multiple influences from the social, psychological, and instructional realms so that their relative influences on attainment can be verified.

This is not meant to imply that statistical tests and inferential analyses can answer all relevant questions. Qualitative data allow us to explore issues that are not easily quantified, e.g., how one's desire to sound native or one's willingness to socialize with native speakers impacts ultimate attainment, and how such factors relate to AO (and to the current age of the participant). All of these factors can be measured and tested through scalar and categorical response-types, and subsequently validated as either significant for accent or not, but only qualitative techniques such as interviews offer an in-depth way to explore how they operate for the individual.

### **METHODOLOGICAL RECOMMENDATIONS**

Based on the discussion above, I outline specific recommendations for participant selection, task design and procedures, ratings, and analyses. All of these suggestions are geared toward enhancing validity within studies, comparability across samples, and overall reliability.

### **Recommendations for Participant Selection**

In studies that aim to test the critical period hypothesis, selection needs to be carefully considered with an eye toward replicability and generalizability. Should aspects of experience and orientation continue to be pre-selected, or should a range of such factors be assured by allowing for a more random sample selection? Furthermore, to what extent can evidence be compared across studies when learner groups are specialized, or exceptional? Conditions and criteria for participant selection should be clearly justified. If the researcher's position is that only 'exceptions' to the critical period are needed in order to refute it (as suggested by Long, 1990), s/he may opt to include only exceptional learners; however, the sample should include a range of ages at first exposure to confirm whether early onset is the primary force behind such exceptional ability (Hyltenstam & Abrahamsson, 2003).

In terms of comparing age effects to other influences, if a study seeks to describe the effects of long-term experience on attainment, then participants' actual experience with the target language must be clearly specified. For example, if a study hopes to contrast age effects with experience, both formal and informal L2 experience types (e.g., instruction, contact with native speakers, etc.) should be carefully documented so that any impact can be measured against that of age. If learner orientation is the focus of investigation, participants with compelling reasons for acquiring the target language, desire to sound native, affiliation with the TL community, etc., should be recruited for study, and these criteria should be sufficiently measured and analyzed against concomitant influences like AO and length of residence.

### **Recommendations for Task Design**

The earlier discussion highlighted concerns about internal validity in task design, and several things can be done to ensure a more solid instrument along these lines. In order to represent various degrees of control and naturalness, a range of tasks should be included. Task design should ensure a spectrum of isolated and contextualized task types to target various levels of phonological processing, including segmental, syllabic, prosodic, and rhythmic abilities. (This principle could equally be applied to perception-based studies, which too often rely only on isolated segments for data points). Such a range not only addresses authenticity concerns, it allows the researcher to test for task effects that may provide insights on isolated versus global-level fluencies.

In addition to the task itself, we should consider the instructions and procedures given to participants. While multiple takes may be appropriate for novice learners, tests of truly advanced learners should not allow practice to obtain good tokens. For participants at this level, open-ended response types are appropriate, and should be considered a standard.

One further suggestion is that future instruments manipulate constructs like topic familiarity and formality within task types to check for any possible impact on production or perception accuracy. By the same token, varying interactional

dynamics can reveal how phonetic inaccuracies are dealt with in context; in other words, how interlocutors actually accommodate certain kinds of phonetic substitutions (see Jenkins, 2000, 2002). This would give us some indication of what accent implies in the broader context of real communication.

### **Recommendations for Raters and Ratings**

Several suggestions could enhance the validity of rater selection, and the standardization of ratings criteria. First, raters should be selected from a range of backgrounds to ensure a balance of sensitivity to various accents (see Thompson, 1991, for relevant findings), and their ratings should be formally (statistically) tested for comparative reliability. In so doing, any language background differences can be verified as inconsequential to ratings (if reliability is poor, some ‘outlier’ raters may need to be eliminated from the analyses). Second, 5-to-6-point rating scales could easily include both categorical and scalar responses to best represent the variety of judgments made, for example, comprehensibility, authenticity, regional versus standard pronunciation, etc., in addition to scalar evaluations of nativeness (see Albrechtson, Henriksen, & Faerch, 1980; Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing & Munro, 1997; Moyer, 2007a). This could capture information on the impact of a non-native accent for the native speaker (NS) listener.

Rater response types should include discrete measures, but also allow for open-ended comments and notes. This recommendation is based on an instrument introduced in Moyer, 2004: (1) a survey asking raters to prioritize various criteria for judging nativeness (e.g., phonological, grammatical, lexical, and pragmatic aspects of fluency); (2) room for written comments while listening. This allowed for a comparison of rater priorities reported *a priori* – before listening – with the actual features noted that contributed to judgments of strong accent. (Phonological criteria were overwhelmingly ranked first *a priori* and figured most prominently in the ‘online’ comments as well.) In addition, comments written during listening can identify lexical and morphological errors, as well as phonological inaccuracies, thus providing some insight into criteria used in rating more complex, global tasks.

### **Recommendations for Factor and Analysis Breadth**

As noted, both quantity and quality of learner experience must be accounted for in any study of age effects, specifically: domains for L2 contact, personal ties with native speakers, access to formal and informal feedback and instruction, reasons for acquiring the TL, and so on. By gathering a broader set of data, we can accomplish three important objectives: (a) statistically validate the contributions of these factors for phonological attainment; (b) verify areas of conflation or overlap between factors that can disguise the underlying nature of the mechanisms at work (such as maturational effects); (c) compare the relative strength of multiple variables that predict the variance in outcome. With the complex relationships between multiple influences in mind, binary categories (such as *instrumental vs. integrative* motivation) and scalar response types can be supplemented by open-

ended questions that allow for more in-depth understandings of learner experience and orientation.

Finally, incorporating inferential tests, such as multiple regression and ANOVA, allows us to test the relative and *independent* strength of multiple variables as they predict the variance in outcome. These tests sometimes provide the (surprising) result that age is not as powerful a predictor of native-like attainment as we commonly assume (see Bialystok & Hakuta, 1999; Moyer, 1999, 2007a; cf. DeKeyser, 2000 – see Long, 2005, for discussion). More sophisticated means of analysis, such as partialling out the effects of potentially conflated variables, are far less common in ultimate attainment studies (see Long, 2005, for brief discussion), but should be considered a standard for validating the effects of age in relation to its concomitant influences.

## CONCLUSIONS

Though no singular neurological change has been identified as responsible for age effects (see Bialystok & Hakuta, 1999), a late start may well indicate declining auditory and articulatory flexibility, and possible declines in memory, attention and/or perceptual acuity. Some have suggested that relevant neurobiological influences do not stand alone, but are connected to affect, and that, together, these impact language processing and attainment (Pulvermüller & Schumann, 1994; Schumann, 1994, 2001). For now, we may not be able to understand why some late learners are better at acquiring a native-like accent than others, but we can explore why some learners actively seek opportunities to practice and refine their skills, ask for overt feedback on their pronunciation, and develop conscious strategies to improve it. This broader focus on the learner as an active participant in the process would go a long way towards confirming the convergence of affective and cognitive influences in terms of actual L2-oriented behaviors.

More than 15 years ago, Scovel (1988) asserted that a ‘sensitive’ period for language learning is shaped “not by nature, but by nurture,” meaning environmentally-shaped forces such as attitudes, learning contexts, and so on (p. 85). Indeed, given the uncertainties surrounding the nature of mechanisms responsible for acquisition, and their apparent decline coinciding with (or even previous to) biological maturation,<sup>9</sup> a more inclusive view of age effects is more tenable (and hopefully as verifiable), as more traditional accounts.

While personal engagement in the target language varies highly across L2 learners, those with an early AO are, statistically speaking, far more likely to accrue instructional experience and greater personal contacts to native speakers (Moyer, 2004). Early onset does not simply predict greater exposure or contact, but a more advantageous balance of exposure types – experiential *quality*, so to speak – positively affecting both affective and cognitive strategies for improving pronunciation abilities (see Moyer, 2004, 2007b). Even while recognizing these important connections between AO and L2 experience, we cannot assume direct

relationships between this experience and ultimate attainment; understanding the age factor requires an appreciation for how developing L2 experience feeds into learner goals, decisions, and behaviors (see Beebe, 1985; Bialystok, 1997; Klein, 1996; Seliger, 1983), which directly and/or indirectly lead to attainment outcomes. One concrete example of significant behavior referred to earlier illustrates this point: Some L2 learners who reside in-country shift over time to predominantly rely on L2, leaving L1 for special contexts (like family domains), and this affects accent to a significant degree (Flege et al., 1999; Yeni-Komshian et al., 2000). Again, the correlation between such a shift and AO has been similarly documented. Singleton (2000) puts it simply as follows:

...Whereas immigrants arriving at ages older than 10 tend to maintain their L1, immigrants arriving before age 10 seem to switch their dominant language from the home language to the language of the host country. The implication of this phenomenon is that some studies purportedly focusing on L2 proficiency may, in fact, be reporting on a language which has effectively become an L1 for the subjects in question (p. 83).

Singleton's point underscores the fact that an early start signifies much more than neurobiological advantage; it predicts how consistently the target language will be used across various domains, and the likelihood that it will eventually become the dominant language.

This paper has argued that age effects research in L2 phonology can do more to enhance validity and reliability in its methodologies, by standardizing ratings procedures and inter-rater reliability measures, broadening task types for advanced learners, and designing elicitation techniques that are appropriate to the participants involved. Following the recommendations outlined above, internal and external validity can be improved to an appreciable extent, and important findings can be more readily replicated, which will hopefully lead to clearer, even generalizable, principles regarding what late learners can and cannot do in the phonological realm. Finally, greater sophistication in statistical analyses will allow the effects of age to be more clearly delineated from other conflated factors – a reasonable and timely goal for critical period studies.

### **ACKNOWLEDGMENTS**

The author would like to express gratitude to Mike Long for his careful reading of an early draft of this paper. His insightful comments and suggestions undoubtedly improved the work, even if he essentially disagrees with the position taken here (and elsewhere) on the substance and relative strength of age effects in SLA! I would also like to thank the anonymous reviewers whose recommendations were greatly appreciated. As always, all opinions, unwitting omissions, and errors are strictly my own.

## NOTES

<sup>1</sup> Internal *validity* – the extent to which the instrument and analyses accurately captures the construct(s) under examination – naturally affects external validity – the ability to generalize the findings based on the potential for others to replicate the same study under parallel conditions.

<sup>2</sup> The actual cut-off point for defining ‘late’ learners is also problematic. Across studies on critical period effects, ‘late’ can be defined as anywhere from age 9 (where ‘early’ is defined as age 8 or below) to age 15, or even to age 20 or later (cf. MacKay & Flege, 2004; MacKay, Flege, Piske, & Schirru, 2001; Yeni-Komshian, Robbins, & Flege, 2001). Grouping participants aged 15 to 26 (and older) may be justified on the grounds that age effects are effectively non-existent (at least from the point of view of statistical significance) past the offset of puberty (around age 15-17 according to Johnson & Newport’s 1989 data). At the same time, the period from 10-14 is not typically treated as interesting, in spite of the fact that this age range may indeed pinpoint the phase at which the linear relationship between age and outcome breaks down, at least for production abilities in L2 phonology. (For this reason, Moyer in her 2004 work further reduces the original 3 age groups in her analysis to 4, in order to illustrate the nature of the linearity within this 9-13 range (pp. 94-95). Some recent studies have contrasted ‘very early’ onset with ‘early,’ i.e., comparing AO of 1-5 to that of 7-8 (Yeni-Komshian, Robbins, & Flege, 2001).

<sup>3</sup> *Reliability* is the stability of a measure from one application to another, i.e., the consistency of responses from one set of items to another similar set, or from one data collection to another subsequent collection using the same instrument. Threats to reliability are always a concern when dealing with specialized abilities (see Porte, 2002, for discussion). Thus, the need to adequately take into account L2 learning contexts and experience for pooled data from late learners is essential; comparability among individual learners within such groups should be of concern.

<sup>4</sup> Considering the research on instructional effects overall, Piske et al. (2001) suggest that total amount of instruction may be far less significant than *type* (e.g., instruction targeted to phonetic accuracy, suprasegmental authenticity, etc.). The authors also point out that pronunciation receives little attention in classrooms these days, and thus it is no surprise to find contradictory results across studies – there is undoubtedly great variation in quantity and approach overall. At a minimum, we can assume that classroom experience provides some opportunity for feedback on both segmental and suprasegmental precision, which may encourage “internal” monitoring (see Dickerson, 1987; Van der Linden, 1995) and self-initiated strategies that improve authenticity (for evidence, see Elliott, 1995b; Moyer, 2004; Osburne, 2003), e.g., mimicking a native speaker model.

<sup>5</sup> This paper will not treat perception tasks, per se, although the question of whether late learners can perceive new phonetic categories more accurately than they can produce them informs a lively debate in the research. Potentially “asymmetrical” abilities along these lines may point to an underlying decline in motor control over articulatory organs, thus affecting accent in spite of one’s possibly having a ‘good ear.’ Based on their recent survey of research on early and late bilingualism, Sebastián-Gallés and Kroll (2003) assert that the auditory perceptual system remains highly plastic at least through early childhood and possibly much later, but beyond early childhood, those who have learned

to differentiate difficult L1-L2 contrasts seem to do so using different acoustic parameters than the earliest (native) listeners (p. 290).

<sup>6</sup> Spectrographic analysis is not subject to the validity concerns raised here because it is not based on impressionistic measures (which can be criticized for being subjective).

<sup>7</sup> In a qualitative comparison of teacher and non-teacher perceptions of L2 speech, Hadden (1991) found that non-teachers were less critical of non-native abilities; however, Thompson (1991) reports that inexperienced raters are more sensitive to foreign accent (cited in Piske et al., 2001). This contradicts Bongaerts et al. (1997) who found significant overlap between experienced and inexperienced raters. Cesar-Lee (2000) found that formant frequencies, syllable duration, and VOT values were the criteria that teachers actually used, though lexical and grammatical errors and the teacher's own level of fluency also correlated to judgments of native/non-nativeness.

<sup>8</sup> Southwood and Flege (1999) examined the validity of a metathetic continuum (equal intervals from high to low) versus a prothetic continuum (continuous, but not integral, i.e., not necessarily linear) and found that NS judges were able to discern accent in terms of equal intervals, thus they contend that this approach is valid, but that a 9-point scale is more in line with the "full range of sensitivity" of the judges.

<sup>9</sup> One problem with assuming neurobiological age as primarily responsible for a decline in language learning abilities is that age is a continuous variable. It would therefore be expected to produce a gradual, steady (negative) relationship to outcome. Birdsong (2005) therefore distinguishes age from *maturation* as a specific event that could impact language acquisition in negative ways. For example, if language learning abilities gradually and continually decline, this would suggest *general age effects*. On the other hand, declines beginning as early as age 6 and continuing through age 15-16, followed by a drop-off in linearity or slope (such as those seen in Birdsong and Molis, 2001, Johnson and Newport, 1989, and Moyer, 2004) indicate an event with a specific "window," i.e., maturation. See Long (1990) for discussion of separate critical periods for different language abilities.

## REFERENCES

- Albrechtsen, D., Henriksen, B., & Faerch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning*, 30, 365-396.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42, 529-555.
- Asher, J., & Garcia, R. (1969). The optimal age to learn a foreign language. *Modern Language Journal*, 53, 334-341.
- Beebe, L. (1985). Input: Choosing the right stuff. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 404-414). Cambridge, MA: Newbury House.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Studies in Second Language Acquisition*, 13, 116-137.
- Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 161-181). Mahwah, NJ: Lawrence Erlbaum.
- Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 68, 706-755.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. Kroll & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic perspectives* (pp. 109-127). Oxford: Oxford University Press.
- Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language*, 44, 235-249.
- Bohn, O., & Flege, J. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14, 131-158.
- Bongaerts, T., Planken, B., & Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 30-50). Clevedon, UK: Multilingual Matters.
- Bongaerts, T., Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the production of foreign language. *Studies in Second Language Acquisition*, 19, 447-465.
- Cesar-Lee, B. (2000). Quantification of accented pronunciation by American-English speakers in French as a foreign language. Unpublished doctoral dissertation. University of Florida.
- Cook, V. (2002). Background to the L2 user. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 1-28). Clevedon: Multilingual Matters.
- Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon: Multilingual Matters.
- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-533.
- DeKeyser, R., & Hall, J. (2005). What does the critical period really mean? In J. Kroll and A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.
- Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 20, 1-16.
- Derwing, T., Munro, M., & Wiebe, G. (1997). Pronunciation instruction for "fossilized" learners: Can it help? *Applied Language Learning*, 8, 217-235.

- Derwing, T., & Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning*, 13, 1-17.
- DeWaele, J. (2002). Individual differences in L2 fluency: The effect of neurobiological correlates. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 221-249). Clevedon: Multilingual Matters.
- Diaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26, 249-273.
- Dickerson, L. (1975). The learner's interlanguage as a system of variable rules. *TESOL Quarterly*, 9, 401-07.
- Dickerson, W. (1987). Explicit rules and the developing interlanguage. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 121-140). Dordrecht, The Netherlands: Foris.
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in L2 learning. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 589-630). Malden, MA: Blackwell.
- Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *Journal of the Acoustical Society of America*, 110, 1606-1618.
- Elliott, A. (1995a). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal*, 79, 530-542.
- Elliott, A. (1995b). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *Modern Language Journal*, 79, 356-371.
- Elliott, A. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, 80, 95-108.
- Fathman, A. (1975). The relationship between age and second language productive ability. *Language Learning*, 25, 245-253.
- Flege, J. (1991). Perception and production: the relevance of phonetic input to L2 phonological learning. In T. Huebner & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 249-289). Amsterdam: John Benjamins.
- Flege, J., & Hillenbrand, J. (1987). Limits on phonetic accuracy in foreign language speech production. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology* (pp. 176-203). Cambridge, MA: Newbury House.
- Flege, J., & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 24, 1-34.
- Flege, J., Munro, M., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125-3134.
- Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78-104.
- Hadden, B. (1991). Teacher and non-teacher perceptions of second-language communication. *Language Learning*, 41, 1-24.
- Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 539-588). Malden, MA: Blackwell.
- Ioup, G., Boustagi, E., El Tigi, M., & Moselle, M. (1994). Re-examining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment.

- Studies in Second Language Acquisition*, 16, 73-98.
- Ishida, S. (2003). Taking pronunciation to task: Exploring methodological issues in maturational constraints research. Unpublished manuscript. University of Hawai'i at Manoa.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23, 83-103.
- Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 39, 215-258.
- Kachru, B., & Nelson, C. (1996). World Englishes. In S. McKay & N. Hornberger (Eds.), *Sociolinguistics and language teaching* (pp. 71-102). Cambridge: Cambridge University Press.
- Klein, W. (1996). Language acquisition at different ages. In D. Magnusson (Ed.), *The lifespan development of individuals: Behavioral, neurobiological, and psychosocial perspectives. A synthesis* (pp. 244-264). Cambridge: Cambridge University Press.
- Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley & Sons.
- Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12, 251-285.
- Long, M. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *IRAL*, 43, 287-317.
- MacKay, I., Flege, J., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, 110, 516-528.
- MacKay, I., & Flege, J. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, 25, 373-396.
- Major, R. (1993). Sociolinguistic factors in loss and acquisition of phonology. In K. Hyldenstam & A. Viberg (Eds.), *Progression and regression in language: Sociocultural, neuropsychological and linguistic perspectives* (pp. 463-478). Cambridge: Cambridge University Press.
- Major, R., & Kim, E. (1999). The similarity differential rate hypothesis. In J. Leather (Ed.), *Phonological issues in language learning* (pp. 151-183). Malden, MA: Blackwell.
- Marinova-Todd, S., Marshall, D., & Snow, C. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, 34, 9-34.
- Markham, D. (1997). Phonetic imitation, accent, and the learner. *Travaux de l'Institut de Linguistique de Lund*, 33, 3-269.
- Missaglia, F. (1999). Contrastive prosody in SLA – An empirical study with adult Italian learners of German. *Proceedings of the 14<sup>th</sup> International Congress of Phonetic Sciences*, Vol. 1, 555-558.
- Moyer, A. (1999). Ultimate attainment in L2 phonology. The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21, 81-108.
- Moyer, A. (2004). *Age, accent and experience in second language acquisition. An integrated approach to critical period inquiry*. Clevedon: Multilingual Matters.
- Moyer, A. (2007a). A closer look at the experience construct in L2 phonology: What role does interaction play? Paper presented at the meeting of the American Association of

- Applied Linguistics* (AAAL), Costa Mesa, California.
- Moyer, A. (2007b). Do language attitudes determine accent? A study of bilinguals in the USA. *Journal of Multilingual & Multicultural Development*, 28, 502-518.
- Moyer, A. (forthcoming). Input as a means to a critical end: Quantity and quality in L2 phonological attainment. In M. Young-Scholten & T. Piske (Eds.), *Input matters*. Clevedon: Multilingual Matters.
- Munro, M., & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. In J. Leather (Ed.), *Phonological issues in language learning* (pp. 285-310). Malden, MA: Blackwell.
- Munro, M., & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech. The role of speaking rate. *Studies in Second Language Acquisition*, 23, 451-468.
- Neufeld, G. (1988). Phonological asymmetry in second-language learning and performance. *Language Learning*, 38, 531-59.
- Olson, L., & Samuels, S. (1982). The relationship between age and accuracy of foreign language pronunciation. In S. Krashen, R. Scarcella & M. Long (Eds.), *Child-adult differences in second language acquisition* (pp. 67-75). Rowley, MA: Newbury House.
- Osburne, A. (2003). Pronunciation strategies of advanced ESOL learners. *IRAL*, 41, 131-143.
- Oyama, S. (1976). A sensitive period for the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research*, 5, 261-283.
- Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30, 449-472.
- Piller, I. (2002). Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics*, 6, 179-206.
- Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.
- Porte, G. (2002). *Appraising research in second language learning*. Amsterdam: John Benjamins.
- Pulvermüller, F., & Schumann, J. (1994). Neurobiological mechanisms of language acquisition. *Language Learning*, 44, 681-734.
- Purcell, E., & Suter, R. (1980.) Predictors of pronunciation accuracy: A re-examination. *Language Learning*, 30, 271-87.
- Sato, C. (1985.) Task variation in interlanguage phonology. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 181-196). Cambridge, MA: Newbury House.
- Schumann, J. (1994). Where is cognition? Emotion and cognition in second language acquisition. *Studies in Second Language Acquisition*, 16, 231-242.
- Schumann, J. (2001). Appraisal psychology, neurobiology, and language. *Annual Review of Applied Linguistics*, 21, 23-42.
- Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House.
- Sebastián-Gallés, N., & Kroll, J. (2003). Phonology in bilingual language processing: Acquisition, perception and production. In N. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production*. Berlin: Mouton de Gruyter.
- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, 24, 209-239.

- Seliger, H. (1983). Learner interaction in the classroom and its effect on language acquisition. In H. Seliger & M. Long (Eds.), *Classroom-oriented research in second language acquisition* (p. 246-267). Cambridge, MA: Newbury House.
- Singleton, D. (2000). Age and second language acquisition. *Annual Review of Applied Linguistics*, 21, 77-89.
- Snow, C., & Hoefnagel-Hoehle, M. (1982). The critical period for language acquisition: Evidence from second language learning. In S. Krashen, R. Scarcella & M. Long (Eds.), *Child-adult differences in second language acquisition* (pp. 84-92). Rowley, MA: Newbury House.
- Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In Ritchie, W. & T. Bhatia, *Handbook of second language acquisition* (pp. 375-409). San Diego: Academic Press.
- Southwood, M., & Flege, J. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics*, 13, 355-449.
- Stokes, J. (2001). Factors in the acquisition of Spanish pronunciation. *ITL Review of Applied Linguistics*, 131-132, 63-84.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, 36, 131-145.
- Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, 26, 233-253.
- Tarone, E. (1982.) Systematicity and attention in interlanguage. *Language Learning*, 32, 69-82.
- Tarone, E. (1987). The phonology of interlanguage. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology* (pp. 70-85). Cambridge, MA: Newbury House.
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, 41, 177-204.
- Van der Linden, E. (1995). Fossilization versus monitor use: A case study. *ITL, Review of Applied Linguistics*, 107, 1-16.
- White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, 12, 233-265.
- Yeni-Komshian, G., Flege, J., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, 3, 131-149.
- Yeni-Komshian, G., Robbins, M., & Flege, J. (2001). Effects of word class differences on L2 pronunciation accuracy. *Applied Psycholinguistics*, 22, 283-299.

**Alene Moyer** is Associate Professor at the University of Maryland, where she teaches sociolinguistics, SLA, and German language and linguistics. Her research focuses on second language phonology, and the acquisition of German as a second language. Her book *Age, Accent and Experience in Second Language Acquisition* appeared with Multilingual Matters in 2004.