Nonnative and second-language speech perception:

Commonalities and complementarities[1]

Catherine T. Best[2]
MARCS Auditory Laboratories
Haskins Laboratories

Michael D. Tyler
MARCS Auditory Laboratories

Correspondence: MARCS Auditory Laboratories, U. Western Sydney, Penrith South DC NSW 1797 Australia. e-mails: c.best@uws.edu.au and m.tyler@uws.edu.au.

Word Count: 10389 (inc endnotes) + references.

Best, C. T., & Tyler, M. D. (in press). Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (eds) *Second language speech learning: The role of language experience in speech perception and production.* Amsterdam: John Benjamins. (Accepted 16 Jan 2006).

Abstract

Language experience systematically constrains perception of speech contrasts that deviate phonologically and/or phonetically from those of the listener's native language. These effects are most dramatic in adults, but begin to emerge in infancy and undergo further development through at least early childhood. The central question addressed here is: How do nonnative speech perception findings bear on phonological and phonetic aspects of second language (L2) perceptual learning? A frequent assumption has been that nonnative speech perception can also account for the relative difficulties that late learners have with specific L2 segments and contrasts. However, evaluation of this assumption must take into account the fact that models of nonnative speech perception such as the Perceptual Assimilation Model (PAM) have focused primarily on *naïve* listeners, whereas models of L2 speech acquisition such as the Speech Learning Model (SLM) have focused on *experienced* listeners. This chapter probes the assumption that L2 perceptual learning is determined by nonnative speech perception principles, by considering the commonalities and complementarities between inexperienced listeners and those learning an L2, as viewed from PAM and SLM. Among the issues examined are how language learning may affect perception of phonetic vs. phonological information, how monolingual vs. multiple language experience may impact perception, and what these may imply for attunement of speech perception to changes in the listener's language environment.

Introduction

No one can step into the same river twice … both river and person are ever-changing

Heraclitus [540-480 BC] (paraphrased)

This metaphor for the continual reshaping of an individual's experience of life events by the convergence between their actively changing environment and their unfolding personal history is familiar as a viewpoint in existential philosophy (cf. Kahn, 1981). Yet it may also provide new perspective on our understanding of how an individual's language environment and their progress in acquiring a native and/or a second language together shape the way they perceive speech. As research has undeniably shown, the way listeners perceive phonetic information depends on their linguistic experience and developmental history. Prior contact with the stimulus language, and position along the trajectory of native or first language (L1) development, converge in some crucial way to shape one's perception of phonetic details and phonological structure in speech. Perception differs in important ways between naïve listeners and those who have experience with the stimulus contrasts as elements of a second language (L2). Experiential influences vary as well with age at onset of L2 acquisition, and/or with other crucial aspects of fluency and usage in both L1 and L2. Thus, to the extent that Heraclitus' metaphor applies to experiential effects on speech perception, the infant and the adult could never truly perceive the same speech in the same way, nor could the L2 learner or bilingual perceive L2 *or* L1 speech in exactly the same way as native monolinguals of either language.

We hope to uncover new insights about these issues by examining commonalities and complementarities between naïve listeners' perception of nonnative speech contrasts and second language (L2) learners' perception of L2 contrasts. The centerpiece of this examination will be a critical comparison of two influential theoretical models: Flege's (1995, 1999, 2002) L2 Speech Learning Model (SLM), and Best's (1993, 1994a, 1994b, 1995; Best, McRoberts, & Sithole, 1988) Perceptual Assimilation Model (PAM) of nonnative speech perception. We will attempt to extend the principles of PAM to L2 speech perception, and to present testable hypotheses as a guide for future research. We focus our comparison here on the two listener groups that have been the focus of those to models – late second language learners and functional monolinguals, respectively. In the concluding section, we offer suggestions that may guide research and theory on developmental and experiential influences on speech perception, including that by fluent bilinguals.

Heraclitus' metaphor provides a useful analogy for the fact that young listeners encounter an ever-changing flow of phonetic information across speakers and communicative contexts as they acquire a native language. That phonetic flux differentiates further if a listener begins establishing an L2, which adds the fluctuation between L1 and L2 speech environments, and mixed (code-switched) speech in bilingual environments. Thus, we will begin this chapter by considering the role of the environment in speech perception for the two groups of listeners that have been the focus of PAM and SLM: naïve nonnative listeners and L2-learning listeners, respectively.

The Role of the Environment in the Development of Speech Perception

From early notions of critical periods, we now understand better the important role that the environment plays in language acquisition across ages. The contribution of age to L2 perceptual learning occurs primarily through interactions with length of residence (LOR), relative usage of L1/L2 (Flege, 1999, 2002; Flege & MacKay, 2004), and relative quantity and quality of input from native L2 speakers (see Flege & Liu, 2001; Jia & Aaronson, 2003; Jia, Strange, Wu, Collado, & Guan, 2005, accepted with revisions).

Not only is the amount of exposure important, but the phonetic properties of the language input to learners appears to dynamically interact with their developmental level and L2 learning status. For example, along with prosodic properties that convey positive emotion and affiliation (e.g., Kitamura & Burnham, 2003; Singh, Morgan, & Best, 2002), infant-directed speech (IDS) also contains hyper-articulated vowels (Andruski, Kuhl, & Hayashi, 1999; Kuhl et al., 1997). The fact that emotional prosody, but *not* vowel hyper-articulation, is observed in speech directed to pet cats and dogs suggests that the latter may reflect a didactic function of language learners' speech input properties (Burnham, Kitamura, & Volmer-Conna, 2002; see also, Englund & Behne, 2005; Liu, Kuhl, & Tsao, 2003). Importantly, this didactic function is not only apparent in IDS: foreigner-directed speech (FDS) also shows vowel hyper-articulation, though not the emotional prosody of IDS (Knoll & Uther, 2004).

While systematic shifts in speech to language learners are important in indicating that the linguistic environment is responsive to the addressee's developmental and linguistic status, our primary concern is with how the listener's status on those two dimensions affects their perception of nonnative speech. The target stimuli in studies of experiential influences on speech perception have classically been nonnative minimal distinctions, that is, single phonetic feature contrasts that are linguistically irrelevant in the listener's native language. Therefore, perception of minimal contrasts will be our main concern.

To focus our comparison of naïve listeners and L2 learners, we conceptualize their characteristics, as studied in speech perception research, as lying somewhere along three separate dimensions: 1) L1 acquisition at onset of L2 learning; 2) ratio of L1/L2 usage on an average daily basis, and; 3) ratio of L1/L2 in the language environment. L2 learners will be defined as being engaged in the process of acquiring an L2 (SLA: second language acquisition). By comparison to research on “fluent bilinguals,” the participants of L2-learner studies generally have commenced learning at older ages (later childhood, adolescence or adulthood), show a wider range of L1/L2 usage, and tend to live in more L2-biased environments.

These idealized dimensions obviously cannot do justice to the complexity of variation and overlap among real-world language users, who range from monolingual to fluently multilingual, and vary along multiple and often nonlinear dimensions. Although ‘fluency’ (proficiency, competency) and ‘balance’ (language dominance) are often used in discussions of L2 learners and especially bilinguals, these terms can be fraught with ambiguities, operational measurement vagaries, and theoretical disputes. We have chosen, instead, to identify our three criterial dimensions in presumably more objective terms. However, operationalizing and measuring even these characteristics is likely to be difficult and often indirect. Our conceptualization also differs in certain respects from that of experts who study other aspects of bilingual and L2 learner language skills (see, e.g., Kroll & de Groot, 2005). These differences are not meant to challenge those experts, but rather to serve as a heuristic for our narrower and more circumscribed topic: the effects of developmental language experience on perception of phonetic and phonological information in speech.

Within this simplified framework, we define nonnative listeners more strictly than has often been done in the past. For us, they are functional monolinguals, i.e., not actively learning or using an L2, and are linguistically naïve to the target language of the test stimuli.[3] By comparison, L2 learners are people who are in the process of *actively learning* an L2 to achieve functional, communicative goals, that is, not merely in a classroom for satisfaction of educational requirements. We will next summarize relevant findings and controversies regarding these groups.

Perception of Speech as a Function of Linguistic Experience

Naïve Nonnative Listeners

For adults who lack active language-learning experience with a non-L1, especially the stimulus language, the pattern of nonnative speech perception is relatively well-established: functional monolinguals have notable difficulty categorizing and discriminating many phonetic contrasts from unfamiliar languages that are not used to distinguish lexical items in their own language, including both nonnative consonants and vowels (e.g., Abramson & Lisker, 1970; Goto, 1971; Lisker & Abramson, 1967; Polka, 1992; Polka & Werker, 1994; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001; Veleva, 1985; Werker & Lalonde, 1988; Werker & Logan, 1985; Werker & Tees, 1984).

More recent work, however, has found that not all nonnative segmental contrasts are equally difficult. Some are discriminated moderately well and others at nearly native-like levels (Beddor & Strange, 1982; Best, Hallé, Bohn, & Faber, 2003; Best, McRoberts, & Goodell, 2001; Best et al., 1988; Best & Strange, 1992; Best, Traill, Carter, Harrison, & Faber, 2003; Kochetov, 2004; Polka, 1991; Polka, 1995; Strange & Bohn, 1998). In addition, the relative ease or difficulty of a given contrast varies according to the listener's native language (Best, Hallé et al., 2003; Best & Strange, 1992; Best, Traill et al., 2003; Flege, 1989; Hallé, Best, & Levitt, 1999). For example, English-speaking adults discriminate click consonant contrasts from southern African languages quite well (Best & Avery, 1999; cf. Best et al., 1988), indeed *better* than listeners of other African click languages for certain nonnative click contrasts (Best, Traill et al., 2003). Variations in discriminability of nonnative speech contrasts extend beyond vowels and consonants, to perception of tonal contrasts from tone languages (Burnham & Francis, 1997; Burnham, 2000; Hallé, Chang, & Best, 2004; So, 2005). These variations across stimulus contrasts and listener languages are generally thought to reflect the perceived similarities/dissimilarities between phonetic properties of the nonnative stimuli and those of the listener's native phonology.

Consistent with that notion, listeners are responsive not only to phonetic details that are of potential phonological relevance to the native language, but also to gradient-like (non-contrastive) phonetic variation *within* consonant or vowel categories. They show systematic contextual influences and within-category organization when rating phonetic variants within a native category as ranging from "good" to "poor" (or "hyper") (Allen & Miller, 2001; Brancazio, Miller, & Paré, 2003; Miller, 1994; Miller, Connine, Schermer, & Kleunder, 1983; Miller & Volaitis, 1989; Volaitis & Miller, 1992). Discrimination of exemplars has also been reported to be asymmetrical, that is, worse among good exemplars of the category than among poor exemplars, which has led to a proposal that language experience causes a "warping" of perceived similarity around the ideal, or prototypical, representative of a known phonetic category (Grieser & Kuhl, 1989; Iverson & Kuhl, 1995, 1996; Kuhl, 1993).[4] Given the evidence of sensitivity to within-category phonetic variations, it should perhaps not be surprising that nonnative speech perception is systematically affected by fine-grained phonetic similarities and dissimilarities between native and nonnative phones (e.g., Best, Faber, & Levitt, 1996; Best, Hallé et al., 2003; Best & Strange, 1992; Hallé et al., 1999), and is not constrained only to potential phonological distinctiveness. Monolingual adults' perception of nonnative phonetic contrasts is also systematically influenced by native phonotactic biases (Flege, 1989; Hallé, Best, & Bachrach, 2003; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kochetov & So, 2005; Strange et al., 2001; Strange & Bohn, 1998; see also Weber, 2001), coarticulatory patterns (Beddor, Harnsberger, & Lindemann, 2002; Bohn & Steinlen, 2003), and allophonic or other phonetic variations (Flege & Hillenbrand, 1987; Gottfried & Beddor, 1988; Harnsberger, 2000, 2001). Relatedly, specific phonetic properties of speaker and/or token variations can impact on listeners' discrimination and categorization of a nonnative phonological contrast (e.g., Kingston, 2003). Finally, discrimination asymmetries for some nonnative contrasts may reflect within-category perceptual differentiation for the closest native consonant(s) or vowel(s) (Iverson, 2005; Iverson et al., 2003; Polka, 1995; Tsukada et al., 2005). However, the asymmetry has sometimes reflected *better* discrimination for the exemplars that are more like the native ideal (Best et al., 2001), or failed to show a clear relationship with perceived similarity to the native category (Flege, Munro, & Fox, 1994). Some asymmetries, moreover, have been observed across both native and nonnative listeners, suggesting universal rather than experience-tuned biases (Polka & Bohn, 2003).

Thus, perception is not confined to differences that are relevant to native phonological contrasts. Adult monolinguals also show systematic perceptual sensitivities to non-contrastive phonetic variation in both native and nonnative speech. With nonnative speech, some aspects of sensitivity to phonetic variation are related to similarities between nonnative stimuli and native speech patterns, but others reflect language-universal perceptual tendencies. The implications of these experience-tuned vs. universal phonetic sensitivities have not yet been fully resolved.

L2-learning listeners

We turn next to adult L2 learners, who are somewhere along in the active process of becoming speakers of a language other than their native one.

Numerous L2 speech researchers have posited that a learner's L1 and L2 phonological systems are not completely separate, but are instead situated within an encompassing *interlanguage.* A key question, then, is whether L2 learning-related perceptual adjustments result in systematic changes in *L1* speech perception, implying a perceptual influence of L2 learning beyond the L2 itself. One can also imagine a more global possibility, that a listener's perception of phonetic information *generally* behaves as a coherent dynamical system in which changes in any subregion may affect the lay of the remaining landscape. That is, contact with a different dialect of the L1 could conceivably cause perceptual changes. Comparisons among L2, native, unfamiliar native (dialects) and nonnative speech perception would help elucidate how phonological and phonetic information is organized within, between, and beyond the listeners' language(s), and would help determine how that organization changes with increasing knowledge and use of another language or dialect, thus elucidating the effects of language experience on speech perception. Unfortunately, there have been almost no investigations on L2 learners' perception of *unfamiliar* nonnative contrasts or of *native* contrasts (none for other dialects). Existing findings are essentially limited to perception of L2 speech contrasts.

The picture appears similar, though not identical, to that for functional monolinguals. Initial findings with late L2 learners suggested difficulties developing perceptual differentiation of L2 contrasts that are not used in their L1. However, those participants had rather limited L2 experience, particularly as produced by native speakers. Subsequent work with a wider range of L2 experience, learning contexts, phonetic contrasts, and specific L1s and L2s has shown that a larger number of factors impinge on perception in L2 learners as compared to monolinguals.

A critical question has been whether and how perceptual learning of L2 phonological and phonetic properties may occur. Numerous laboratory-based training studies have addressed this. While training studies are of potential interest for understanding L2 learning in fairly controlled situations such as classrooms, our interest lies instead with *natural* communicative situations, which more broadly engage the multi-tiered grammatical and phonological structure of the L2. This is more consistent with theories of second language acquisition (SLA) that hypothesize meaningful conversation is the main context in which the properties of a new language are learned (e.g., Carroll, 1999).

The distinction between learning an L2 in natural communicative contexts vs. in more constrained contexts also arises in comparing SLA (immersion contexts) and FLA (classroom foreign language acquisition, see also Piske, this volume). In many respects, FLA is notably less than ideal with respect to the natural ecology of language learning: It usually occurs in a pervasive L1 setting and does not extend much outside the classroom. It often employs formal instruction on lexical and grammatical information to a much greater extent than in live conversation. When spoken in the classroom, the L2 is often uttered by L1-accented teachers or, at best, by speakers from diverse L2 varieties, thus presenting a variable (or incorrect) model of L2 phonetic details. Indeed, dialect differences can interfere with perception even for native listeners of the L2 (Bundgaard-Nielsen & Bohn, 2004). Thus FLA is a fairly impoverished context for L2 learning, and perceptual findings for FLA listeners should not be conflated with those for L2 listeners (SLA). Nonetheless, FLA conditions do provide a potentially useful (though not tightly controlled) basis of comparison to SLA. FLA listeners, like L2 listeners but unlike monolinguals, have exposure to the target language. Yet unlike L2 listeners or monolinguals, FLA listeners have L2 exposure primarily through *formal instruction* in a *restricted setting*, with *little or unsystematic conversational experience* with *native speakers*. These issues could serve as an important focus for future research. However, we restrict the following discussion to SLA, the context of L2 learning that has been the focus of SLM.

For SLA in adulthood, initial reports found that perception was fairly poor for L2 consonant contrasts that do not occur in their L1s, similar to early studies of monolinguals (e.g., /r/-/l/ for Japanese SL-American English: MacKain, Best, & Strange, 1981). Also, as noted for naïve nonnative listeners, more recent SLA perception findings indicate that categorization and discrimination performance levels vary across L2 contrasts and across L1s, in ways that are systematically related to the contrastive phonological and gradient phonetic properties of the L1s (/r/-/l/, /b/-/v/ vs. /f/-/v/ for Japanese vs. Mandarin SL-Canadian English: Brown, 1998; /w/-/v/ for Sinhalese vs. German SL-British English: Iverson, 2005), and even between L1 dialects (/f/-/θ/ for Quebec French vs. European French [vs. Japanese] SL-Canadian English learners: Brannen, 2002). Again in line with monolinguals, SL learners' perception of L2 contrasts varies systematically according to L1 phonotactic, allophonic, and coarticulatory patterning (e.g., word-final /t/-/d/ for Mandarin vs. Shanghainese vs. Taiwanese SL-American English: Flege, 1989).

Much research on SLA adults' perception of L2 contrasts has focused on vowels. Vowels are important to examine for effects of perceptual experience because they differ in many respects from consonants, both physically (e.g., acoustic and articulatory properties) and linguistically. For example, vowels place fewer constraints on lexical selection than do consonants, even across languages that differ substantially in the size and distinctiveness of their vowel systems, and in the way their vowels are affected by the language's rhythmic characteristics (syllable- vs. stress- vs. mora-timing, e.g., Cutler, Sebastián-Gallés, Soler-Vilageliu, & van Ooijen, 2000). Yet despite those phonemic class differences, findings on SLA adults' perception of L2 vowels in large part mirror the patterns found with L2 consonants.

Native and SLA listeners often differ as to which stimulus dimensions most strongly influence their perceptual differentiation of the L2's vowels (Bohn & Flege, 1990; Fox, Flege, & Munro, 1995) and consonants (Iverson, 2005; Iverson et al., 2003). Moreover, not only can different L1s have different effects on perception of the same L2 vowels, but in addition, L1 and L2 dialect differences can both systematically affect perception of L2 vowel contrasts (Escudero & Boersma, 2004; Ingram & Park, 1997).

A core question has been whether SLA listeners show perceptual learning of L2 contrasts that are initially difficult to differentiate. The findings suggest that the answer is a qualified 'yes.' Perceptual learning occurs for some L2 contrasts, but seems to depend on their phonological and phonetic relationship to the L1, specifically on perceived similarities vs. dissimilarities to L1 phonemes. In addition, some evidence suggests that perceptual learning is influenced by familiarity with the L2. Much of the evidence comes from comparisons between learners who are relatively inexperienced vs. more experienced with conversational speech in an L2-speaking country. More experienced listeners categorize and discriminate certain nonnative L2 contrasts significantly better than less experienced listeners, but generally less well than native L2 speakers (e.g., L1-Japanese and L1-Arabic learners of American English: Best & Strange, 1992; Flege, 1984; MacKain et al., 1981), except in subjectively familiar words (Flege, Takagi, & Mann, 1996). By comparison, for other L2 phonemes and contrasts that are less difficult for monolinguals and fairly inexperienced L2 listeners, identification and discrimination is not dramatically affected by level of experience.

L2 vowels also show evidence of systematic variations in perceptual learning. Vowels that are perceived to differ notably from any L1 vowels yield larger perceptual learning differences than those perceived to be identical or similar to some L1 vowel(s). Also, the perceptual dimensions along which more experienced listeners distinguish these vowels are more similar to those of native listeners, that is, L2-appropriate vowel perception improves with experience and ability. Moreover, perceptual skill level is positively correlated with accuracy in producing the L2 vowels (German, Spanish, Mandarin, Korean, Japanese SL-American English learners: Bohn & Flege, 1990; Flege, Bohn, & Jang, 1997; Fox et al., 1995; Korean vs. Japanese SL-Australian English learners: Ingram & Park, 1998; Yoneyama, 1997), and with ratio of L2/L1 usage (Italian SL-American English learners: Flege & MacKay, 2004; Flege, MacKay, & Meador, 1999). It is important to note, though, that greater L2 usage and proficiency are associated not only with increased L2 production experience, but also with increased L2 listening experience in meaningful conversation.

An important caveat must be noted, however: the definitions of “experienced” vs. “inexperienced” L2 learners have not been standardized, and vary widely from study to study.[5] Perceptual benefits with experience have been reported in some studies for listeners who have lived for as little as 6 months in an L2-speaking country when compared to listeners with even less or no such experience (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; MacKain et al., 1981). Other studies, though, have defined the cut-off for “experienced” as 2, 3, 5 or even 10+ years in an L2 environment. We would argue that none of these later cut-offs is likely to be optimally informative, particularly if “inexperienced” is correspondingly defined as less than 2, 3, 5, or especially 10 years in the L2 environment. We suspect, instead, that the “experienced” cut-off should be set fairly low (e.g., at 6-12 months of experience), for the following reasons. First, significant L2 perceptual learning has been observed in late learners after as little as 6-12 months of SL immersion, as compared to those with 0-6 months of experience. Second, very little perceptual benefit seems to accrue from additional experience past that initial period for most late learners (see, e.g., Flege & Liu, 2001; Jia et al., 2005, accepted with revisions)[6]: No significant differences have been found between SLA adults with 6-12 months of experience and those with 1.5 years, or between those with +/- 2 years, or +/- 5 years of experience (see Aoyama et al., 2004; Jia et al., 2005, accepted with revisions; Tsukada et al., 2005).

Thus, the bulk of perceptual learning may actually take place fairly early in late-onset SLA.[7] It would make both intuitive and theoretical sense for perceptual adjustment to take place early in acquiring a language, before the lexicon and higher-order linguistic structures (morphology, syntax) are well-established. Higher-order structure and meaning, including lexical organization, may somehow help to establish the language-specific relationship between phonetic details and phonological structure, whether in the L1 or an L2. A number of researchers have posited that in L1 acquisition it is the rapid expansion of the lexicon, which coincides with the emergence of morphology and the beginnings of syntax for most toddlers, that may serve as the primary impetus for lexical phonology to emerge from earlier perceptual attunement to language-specific phonetic properties of the native language.

Theoretical models: Extending Nonnative Perception to Meet L2 Perception

Nonnative speech perception and L2 speech perception have frequently been assumed to be essentially identical, that is, to reflect the same L1 influences. Theoretical models of each have sometimes been cited as though they are interchangeable, or as though each approach is intended to account for both naïve monolinguals and L2 learners. But the links and/or similarities between the two theoretical foci have not, in fact, been systematically examined, either theoretically or empirically. The two models that are most frequently cited this way are the Perceptual Assimilation Model (PAM: e.g., Best, 1994a, 1994b, 1995) and the Speech Learning Model (SLM: e.g., Flege, 1995, 1999, 2002). PAM was developed specifically to explain nonnative speech perception by naïve listeners, whereas SLM was designed to address production of L2 speech by L2 learners. Neither model was developed to address both situations, although each has been cited as though it had been. Our aim is to probe the commonalities and complementarities of the two models, and especially to explore whether and how SLM can be used as a starting point to extend PAM's nonnative speech perception framework to L2 learners.

A few additional caveats should be noted first. Whereas both PAM and SLM have sometimes been characterized as being based solely on L1 phonological distinctions (i.e., conceptually comparable to the Contrastive Analysis Hypothesis: CAH), and/or on L1 interference (i.e., phonological suppression), these characterizations are misrepresentative. Neither PAM nor SLM restrict their predictions about native language influences solely to the influence of phonological contrasts in the L1. Both have considered at length, and in numerous publications, the importance of non-contrastive *phonetic* similarities and dissimilarities between L1 and nonnative/L2 phones, including notions of phonetic goodness of fit, and the relationship between phonetic details and phonological categories and contrasts. Flege, in particular, has rejected the CAH account of L2 speech learning in his discussions of SLM. PAM, for its part, rather than invoking phonological interference, has characterized the nature of L1 influence as being based on perceptual learning of phonetic-articulatory patterning at both the abstract contrastive level and, importantly, at the level of non-contrastive gradient phonetic detail. Thus, evidence favoring allophonic or phonotactic or sub-phonemic (phonetic) influences of the L1 on perception of L2 or nonnative speech can be quite consistent with the principles of PAM (e.g., Best & Strange, 1992; Hallé et al., 1999) and/or SLM (e.g., Flege, 1989; Flege & Hillenbrand, 1987).

The Perceptual Assimilation Model of Naïve Nonnative Speech Perception

A core aim of this chapter is to extend PAM to address the issues of L2 learners' perceptual difficulties and biases across variations in target languages and contrasts. This extension, like PAM itself, is founded on the ecological direct-realist premise that the focus of speech perception is on information about the distal articulatory events that produced the speech signal (e.g., Best, 1994a, 1994b, 1995). That metatheoretical assumption is, in turn, compatible with the principles of Articulatory Phonology, which posits that the dynamic parameters of articulatory gestures in speech production serve as the primitives for (language-specific) phonology (e.g., Browman & Goldstein, 1989, 1990, 1992, 1995). Whereas the traditional subject groups for PAM and SLM are different, the two theoretical approaches overlap or diverge in other ways. We will also examine whether and how the L2 extension of PAM may converge or contrast with the principles and predictions of SLM.

Both SLM and PAM have been described in detail in previous theoretical and empirical reports, so we will summarize them just briefly. PAM posits that, when listening to an unfamiliar nonnative phone (phonetic segment), naïve listeners are likely, due to their native language experience, to perceptually *assimilate* the nonnative phone to the most articulatorily-similar native phoneme. That is, it will be heard as a good or even a poor exemplar of a native phonological segment (*Categorized*), or as unlike any single native phoneme (*Uncategorized*) or, rarely, as a non-linguistic nonspeech sound (*Non-Assimilated*).

To predict how naïve listeners will identify and discriminate nonnative phonological contrasts, PAM takes into account how each phone in a contrasting nonnative pair is perceptually assimilated. Each pattern of contrast assimilation entails expectations about discrimination performance. Very good to excellent discrimination is predicted for Two Category (TC) assimilation, in which the two nonnative phones are perceived as acceptable exemplars of two different native phonemes. Poor discrimination is predicted for Single Category (SC) assimilation, in which the two nonnative phones are judged as equally good or poor tokens of the same native phoneme. Discrimination is intermediate for a Category Goodness (CG) difference, in which both of the contrasting nonnative phones are heard as tokens of a single native phoneme, but they differ in goodness of fit to that phoneme. These cases, which all involve categorization of the nonnative phones to native phoneme(s), differ from those in which one or both nonnative phones fail to sufficiently match any given native phoneme, that is, one or both are Uncategorized. One nonnative phone may be perceived as a native phoneme while the other is heard as an uncategorized speech sound, i.e., (Uncategorized-Categorized assimilation). It should be discriminated very well because it reflects a phonological distinction between an exemplar of a known phoneme versus something that is not an exemplar of that phoneme. Alternatively, both nonnative phones may be Uncategorized (an Uncategorized-Uncategorized assimilation), which may be discriminated poorly to moderately well depending on the proximity of the two phones to the same or different sets of partially-similar native phonemes. Finally, the rarest case is when both nonnative phones are so deviant from the articulatory properties of native phonemes that they are not perceived as speech sounds at all: both are Non-Assimilable. In these cases, discrimination may range from good to excellent, depending on degree of perceived similarity of the two phones as nonspeech events.[8]

PAM considered both phonological and phonetic levels in explaining how the L1 system constrains the perception of nonnative phones that are completely unfamiliar to the listener. Naïve perceivers are unaware of which phonetic distinctions constitute phonological differences in the unfamiliar target language, and could not possibly differentiate phonetic and phonological levels in nonnative stimuli. They can only recognize phonological distinctions in their own language, and phonetic deviations of the unfamiliar phones from their L1 phonemes. For these listeners, the phonetic and phonological levels are related only in the L1, in which perceived differences at the phonetic level became systematically related to the functional linguistic categories of a phonological system during early lexical and grammatical development. We argue that the phonological level is central to the perception of L2 speech by SL learners, who are developing an L2 (or interlanguage) system, in a way that it cannot be for L2-naïve listeners perceiving unfamiliar nonnative speech. Below, after comparing and contrasting the theoretical underpinnings of SLM and PAM, we offer some hypotheses about how the dynamic interaction between L1 and L2, at both phonetic and phonological levels, might affect the likelihood of learning to perceive L2 phones as distinct from L1 phonemes, and the concomitant effect on L1 phonetic categories and contrasts.

SLM Postulates from a PAM Perspective

In this section we review the central postulates of SLM, highlighting areas where the metatheoretical assumptions of SLM and PAM diverge or converge. We then outline how findings from PAM studies with naïve listeners might complement existing SLM findings on L2 learners. From there, we generate novel testable hypotheses for future research on listeners who are developing experience with nonnative contrasts as aspects of a language they are actively acquiring. To frame this discussion, we present below the original SLM postulates (P1 to P4) from Flege (1995) and evaluate each from a PAM perspective. We hope that this comparative discussion of PAM and SLM will offer useful insights for research on L2 speech perception.

P1: The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning. Although SLM is not framed in terms of direct realism, the notion that perceptual learning processes remain intact throughout life is compatible with PAM's direct realist perspective, most specifically with the classic perceptual learning principles of the ecological theory of perception (Gibson & Gibson, 1955). Throughout the lifespan, perceivers continue to refine their perception of speech even for their native language. Changes in the ambient language environment, for example, dialectal innovation or contact with a previously unfamiliar dialect of the L1, may require perceptual learning of new higher-order invariants over age. Learning a different language is a functional extension of this. However, this is not to say that adults learn new speech information, in either their L1 or L2, in exactly the same way as children exposed to the same information. Linguistic and communicative knowledge evolves during an individual's perceptual history. The environment also changes, including the response of others to the individual's appearance and behavior as a physical, cognitive and social being, and particularly as a language-user. For these reasons, an adult is a different sort of L2 learner than a child, not merely for biological or maturational reasons. We are reminded again of the quote from Heraclitus that the same person can never enter the same river twice – both the individual and the waters are inevitably and continually renewed and changed.

Whereas PAM agrees with SLM that the same basic perceptual learning abilities are available to adults learning an L2 as to children learning an L1 or L2, however, it differs from

SLM on other assumptions entailed in P1. Most importantly, PAM posits that perceivers extract invariants about *articulatory gestures* from the speech signal, rather than forming categories from acoustic-phonetic cues. This theoretical distinction is central to discussing SLM’s P2.

P2: Language-specific aspects of speech sounds are specified in long-term memory representations called phonetic categories. PAM’s ecological approach rejects the assumption of mental representation that underlies P2. Rather, the listener directly perceives the articulatory gestures of the speaker and, through perceptual learning, comes to detect higher-order articulatory invariants in speech stimuli. As Gibson and Gibson (1955) argued, expert perceivers have not developed abstract “categories” in long-term memory, but instead have become *tuned* as special purpose devices for perceiving invariants across instances of specific types of objects and events (actions). Similarly, we argue that language users become special purpose devices for perceiving transformational invariants in the amplitudes and phasings of speakers’ vocal tract gestures in the L1 and/or L2. Mental representations of phonetic categories are not required for L2 perceptual learning, from this perspective.

An important implication of the PAM-L2 view is that the perceptual objects/events of interest depend upon the perceiver’s perceptual goals or focus of attention. Some tasks may require attentional focus at a gestural level: the expert listener does not need to detect very fine-grained stimulus details to discriminate [p] from a choking sound. But other contexts of speech perception might require focus at a phonetic level, or at a phonological level. We suggest that perceptual objects/events that are relevant to L2 speech learning are not merely phonetic. Language-relevant speech properties are differentiated not only at the phonetic level but also at the higher-order phonological level, as well as at the lower-order gestural level. L1-L2 differences at a gestural, phonetic, or phonological level may each influence the L2 learner's discrimination abilities, separately or together, depending on the context or the perceiver's goals.

These levels of attentional focus, and the relationship between the L1 and the L2 at each, are essential to explaining L1 and L2 perceptual learning from common theoretical principles. For the remainder of this chapter, we will focus on the phonetic and phonological levels, where the linguistic-functional relationships within the sound system of a language are determined. We use the term *phonological category* heuristically to indicate speech information that is relevant to minimal lexical differences in a given language. *Phonetic category*, on the other hand, is used heuristically to refer to invariant gestural relationships that are sub-lexical yet still systematic and potentially perceptible to attuned listeners, for example, positional allophones, or differing realizations of a given phonological category across dialects or languages. Such phonetic differences do not signal lexical distinctions, but may instead provide perceptual information about the speaker's identity, or their region or language of origin.

P3: Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category. This SLM postulate relies on the theoretical claim made in P2 that speech sounds are specified as phonetic categories, a conceptualization with which PAM disagrees. That metatheoretical issue aside, however, PAM's viewpoint is compatible with the essence of P3. To put it in PAM terms, perceivers continue to refine their perception of speech gestures throughout the lifespan (as in P1). Exposure to L2 phones that are assimilated to an L1 phonological category may require the perceiver to discover a different set of invariants to encompass the new shared phonological category.

It is important to note that one issue, fundamental to this postulate, has not yet received adequate treatment in any model of nonnative or L2 speech perception: How listeners identify nonnative phones as equivalent to L1 phones, and the level(s) at which this occurs. PAM's assumptions differ from those of other models in an important way that would affect how this question is addressed. Other models, including SLM, implicitly take the process to involve *passive* reception of intrinsically *meaningless* *proximal* *stimulus* details (acoustic features) and computation of their statistical distribution in the input, whereas PAM adheres to the direct-realist stance that perceivers *actively* seek intrinsically *meaningful* *distal* *event* information. In speech, this distal information refers to vocal tract gestures and their coordination. Thus, we posit that L1/L2 phones are identified as realizations of the same inter-language phonological category when they are recognized (correctly or incorrectly) as involving functionally the same gestural constellation, for which the parametric details (gestural phasing, constriction location and degree) may or may not differ in a phonetically gradient, rather than phonologically functional, way. That is, contrasts at the functional linguistic level of the L1 phonology and their relationship to phonological contrasts in the L2 are as important to perceptual learning as phonetic categories in the two languages according to PAM but not SLM. In the PAM-L2 framework, it is at the phonological level that listeners may identify L1 and L2 sounds as functionally equivalent (assimilated phonologically). Importantly, such phonological assimilation need not imply that the associated phones are perceived as identical at the phonetic level (e.g., French vs English /r/, see below).

P4: Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space. PAM-L2 would agree with SLM that L1 and L2 phonological categories exist in a common space. However, PAM diverges from SLM in its premise that both phonetic *and* phonological levels interact in L2 speech learning, and that this depends crucially on the relationship between the phonological spaces of the L1 and L2. For example, if an L2 phonological category has been perceptually assimilated (PAM) or perceived as equivalent (SLM) to an L1 phoneme, and if the L1 and L2 phonetic versions of this “in-common” category are discriminable, then we posit that the listener should be able to maintain the L1 and L2 phones as separate phonetic realizations of the one phonological category. The listener could become increasingly perceptually attuned to this phonetic distinction within the single phonological category, and this may cause a shift in the exact details of the L1 and L2 versions of the phoneme relative to those of monolinguals of either language. Thus, perceptual changes may arise as a result of having two phonetic categories for a single interlanguage phonological category. However, as we suggested in our review of L2 speech perception findings, the time window for this L2 learning may be brief and occur early in L2 acquisition; it may possibly be curtailed by increases in learning of higher-order aspects of the L2, such as an expanding lexicon and the acquisition of morphological and syntactic structure. In other words, the focus of attention and learning may shift away from the phonetic level as the learner focuses increasingly on higher levels of linguistic structure.

Another possible type of L2 phonetic category differentiation might result from contrasts at the phonological level. Consider the case of an English L1 listener learning French as an L2, who may develop separate English and French phonetic categories for each of /p/ and /b/, which contrast phonologically in both languages. However, there is a mismatch between the phonological and phonetic properties of stop voicing contrasts in these two languages. French short-lag unaspirated [p] as the phonetic category for the phoneme /p/ overlaps phonetically with the English phoneme /b/, which is often realized phonetically as short-lag unaspirated [p] word-initially. It is specifically at these regions of conflict or ambiguity between language-specific phonetic categories and interlanguage phonological contrasts that we should expect bilinguals to strive to maintain a systematic difference between their overlapping L1 and L2 phonetic categories. That is, we propose that dissimilation may occur between French [p] and English [p], and perhaps between the two interlanguage phonemes /p/ and /b/, which might cause a global shift of all related L1 and L2 categories at the phonetic level, for example, on dimensions other than the primary voice onset time differences.

This example illustrates a range of possibilities that open up if we examine both the phonological and phonetic levels in L2 perceptual learning. In the next section we consider more systematically the possible L1-L2 relationships at the phonological level, and predict the likelihood of success of perceptually learning an L2 phonetic category in each case.

Interactions between L1 and L2 phonological systems, and L2 perceptual learning: PAM-L2

The following examples assume, for simplicity's sake, that the learner begins acquiring the L2 as a functional monolingual, that is, as the type of listener that the original PAM focused on. This type of perceiver has developed an L1 phonological system and L1-specific phonetic attunement, as a result of monolingual language acquisition. PAM's assimilation types describe the possible outcomes of first contact with an unfamiliar phonological system and phonetic patterning. From that point on, for an L2 learner, the goal is to learn the higher order invariants of the L2. Assuming a common L1-L2 system will emerge which incorporates phonetic and phonological levels, our goal is to outline how that system changes over the course of L2 development.

The first question is whether or not the learner has perceived equivalence between an L2 and an L1 phonological category, that is, has perceptually assimilated the L2 phone to this L1 phonological entity. This is compatible with Flege's equivalence classification (e.g., Flege, 1987). However, as we noted above, PAM-L2 addresses equivalence not only at the phonetic level addressed by SLM, but also at the phonological level. Equivalence at the lexical-functional level means that the phonological category has a similar contrastive relationship to surrounding categories in the phonological space. It does *not* automatically imply equivalence or even perceived similarity at the phonetic level. For example, the French /r/, when it is pronounced as a voiceless uvular fricative, bears relatively little phonetic similarity to the English liquid /r/, yet English L2 learners of French tend to equate the lexical-functional category /r/ across the two languages. The fact that the two phonemes are represented identically in the orthographies of both languages may contribute to that bias, but the orthographic commonality may itself reflect the similar patterning of rhotics across the two languages in terms of syllable structure, phonotactic regularities, allophonic and morphophonemic alternations (Ladefoged & Maddieson, 1996; Lindau, 1985). What is important is that English L2 learners of French recognize French /r/ as phonologically equivalent to English /r/ despite clear, perceptible differences between their phonetic realizations. The crucial point we wish to make is that different phonetic realizations may be learned for the phonological /r/ of each language. That is, the L2 learner of French could learn "English" [ɹ] versus "French" [ʁ] phonetic versions of /r/. Moreover, perceptual learning of this type is not limited to the L2 learning situation. The same sort of phonetic differentiation could occur within the L1 for allophones (i.e., non-lexically contrastive phonetic differences such as aspirated and unaspirated /p/ in English /#pV…/ vs. /#spV…/: Whalen, Best, & Irwin, 1997, cf. Pegg & Werker, 1997), or for dialectally differing realizations of phones in contexts of high inter-group contact, such as in most large English-speaking cities (e.g., realization of the interdental fricative /ð/ as a stop [d] in some New York dialects).

To demonstrate how PAM's framework could be extended to predict success at L2 perceptual learning, we will elaborate on four possible cases of L2 minimal contrasts that L2 learners initially perceive as speech segments:

1) Only one L2 phonological category is perceived as equivalent (perceptually assimilated) to a given L1 phonological category. At the phonetic level, if only one member of the L2 contrast is perceived as a good exemplar of a given L1 category, then no further perceptual learning is likely to occur for it. All contrasts with other L2 categories would be either two-category assimilations or uncategorized-categorized assimilations, thus the learner would have little difficulty discriminating minimally contrasting words for those distinctions. In this case we would predict not only that the learner has perceived an L1 and an L2 phonological category as equivalent, but also that the L1 and L2 phonetic categories are perceived as equivalent. To the extent that perceivers continue to fine-tune their perception of higher-order invariants across the lifespan, the common L1/L2 phonetic category could theoretically shift from its previous monolingual setting as the learner accommodates to the new language environment. However, given the perceived phonetic equivalence of the L2 phone as a “good” L1 exemplar, the likelihood of such a shift would seem rather low, or at least small in magnitude. These sorts of questions could be tested with categorical discrimination of the L1-L2 phone pair (see, e.g., Guion, Flege, Akahane-Yamada, & Pruitt, 2000), using multiple speakers of each language.

An alternative possibility is that one L2 phone is perceived as deviant with respect to the closest corresponding L1 phonetic category. That is, the learner perceives that the L2 sound behaves in the same way as the L1 sound at the phonological (functional, lexical) level, but nevertheless perceives a notable phonetic difference between them. The learning of French /r/ by native English listeners would fit this situation. Phonologically and phonotactically, the French /r/ behaves very similarly to English /r/, therefore learners may equate them phonologically. But phonetically, they are quite deviant from one another and should be easily dissimilated at the phonetic level, with French [ʁ] forming a second phonetic category under the common phonological category, /r/. Again, this could be examined with categorical discrimination of the cross-language contrast.

2) Both L2 phonological categories are perceived as equivalent to the same L1 phonological category, but one is perceived as being more deviant than the other. In PAM terms, this would constitute a category goodness assimilation contrast. We would expect learners to be able to discriminate these L2 phones well, though not as well as two category assimilation types. The perceiver should also be able to fairly easily recognize the lexical-functional differences between these L2 phones in minimal lexical contrasts. Therefore, we would predict that a new L2 phonetic *and* phonological category is reasonably likely to be formed eventually for the deviant L2 phone, while the L2 phone that is perceived as a better exemplar would be perceived as phonologically *and phonetically* equivalent to the L1 category. No new category is likely to be learned for the latter. The exact developmental progression for the deviant phone remains a topic for future research. We speculate that, initially, it would be perceptually learned as a new L2 phonetic variant of the L1 phonological category. With continued exposure, the learner should learn to perceive the lexical-functional contrasts between the L2 phones, and to develop a new phonological category for the phonetically “deviant” phone.

The likelihood of the learner developing a new phonetic category for the “better-fitting” L2 phone depends upon the degree of its perceived similarity to the L1 category. Two possibilities arise: a good exemplar will be assimilated to a common L1/L2 phonetic category, with little further perceptual differentiation from the L1 entity, whereas a new phonetic category is likely to be perceptually learned for a less-good exemplar of the L1 category.

3) Both L2 phonological categories are perceived as equivalent to the same L1 phonological category, but as equally good or poor instances of that category. This situation describes a case of single-category L2 contrast assimilation. The learner will initially have trouble discriminating these L2 phones, which would be assimilated both phonetically and phonologically to the single L1 category, and minimally contrasting L2 words would be perceived as homophones. In SLM terms, both L2 phones would be merged with the L1 phonetic category (e.g., Flege, 1995; Flege, Schirru, & MacKay, 2003). Whether or not L2 listeners can learn to perceive a difference between single-category assimilated L2 phones may depend on whether they are perceived as good or as poorer exemplars of the L1 phoneme. However, we would not expect most learners to be able to do so very well. To the extent that a learner *could* perceptually attune to this type of L2 contrast, we hypothesize that they would first have to perceptually learn a new phonetic category for at least one of the L2 phones before they could establish a new phonological category or categories.

A key factor that might influence the likelihood of perceptually learning a new phonological category for this assimilation type is the adaptive significance of detecting the difference between minimally contrasting L2 words. If they are high frequency words, or come from two dense phonological neighborhoods which contain many minimally contrasting words, and if many of these words need to be discriminated for adequate interaction with the surrounding cultural environment, these factors would increase the communicatively relevant pressure to perceptually learn the distinction. If, on the other hand, most contrasting words are low frequency, or from sparse phonological neighborhoods in which there are no minimal-pair words, then the homophony of single-category assimilated contrasts would not greatly disadvantage the L2 learner. This reasoning appears compatible with both PAM and SLM.

4) No L1-L2 phonological assimilation. If the naïve listener does not perceive either of the contrasting L2 phones as belonging clearly to any single L1 phonological category, but rather as each having a mixture of more modest similarities to several L1 phonological categories (Uncategorized, in PAM terms), then one or two new L2 phonological categories may be relatively easy to learn perceptually. This suggestion may appear similar to the SLM concept of the *new* phone, but it differs in some key respects. In PAM’s formulation, it is not only the similarity or dissimilarity of a given L2 phone to the closest individual L1 phonetic category that is crucial to perceptual learning, but its comparative relationships within the interlanguage phonological system. That is, it will also be affected by any other L1 phones that are perceived as similar, and the overlap between those and the ones perceived as similar to the contrasting L2 phone. If each of these uncategorized L2 phones has similarities to *different* sets of L1 phones, that is, they are relatively distant from one another within L1 phonological space, then the listener should easily recognize relevant L2 lexical-functional differences, and two new L2 phonological categories should be perceptually learned. However, if the uncategorized L2 phones are perceived as similar to the same set of L1 phonemes, that is, are *close* to each other in phonological space, then the listener should find them difficult to discriminate, and should not easily perceive relevant L2 lexical-functional differences. In this case, a single new phonological category would be learned, encompassing the two L2 phones. In fact, if there are more than two contrasting L2 phones that are positioned close to one another among the same L1 categories, they may all converge into a single new but undifferentiated phonetic and phonological category. Over the course of L2 learning this category could theoretically split into contrasting L2 categories (e.g., under the sorts of lexical pressures described in 3) above), but it could also remain intact. In the latter case, minimally contrasting L2 words would remain homophonous for the L2 listener.

Learning of phones outside of the phonological space

In the four preceding cases, we considered only L2 phones that are perceived as speech, that is, that lie within the L1 phonological space of the learner. Best, McRoberts and Sithole (1988) examined perception of nonnative consonants that were found, on the other hand, to be non-assimilable to L1 phonological categories. American English monolinguals perceived the isiZulu click consonants as nonspeech sounds, and discriminated the minimal contrasts very well. A further study found that, although isiZulu and Sesotho languages have click consonants in their inventories, native listeners of these languages discriminated certain nonnative !Xõo click contrasts *worse* than American English monolinguals, despite their obviously greater experience with clicks as speech sounds (Best, Traill et al., 2003). In the case of the click contrast on which Americans performed significantly better than isiZulu and

Sesotho listeners, the two African language groups had assimilated the nonnative clicks as equally good exemplars of a single L1 click category.

We know of no studies of L2 perceptual learning of clicks by speakers of non-click L1s, who discriminate clicks well, it would seem, precisely because they tend to hear them as nonspeech sounds, that is, as lacking phonetic or phonological significance. Can speech segments like these that lie outside the listener's L1 phonological space altogether ever become perceptually integrated into that space as speech categories? We do not yet know the answer. Non-assimilated sounds might eventually be perceptually incorporated into the phonological space of the L2 listener as uncategorized speech sounds, possibly resulting in the perceptual learning of 1-2 new phonological category(s). Alternatively, L2 learners may never incorporate these sounds into their phonological space. In this case, these L2 phones could effectively continue to be ignored as nonspeech sounds in linguistic tasks such as word recognition, which L2 listeners could accomplish based only on the remaining phonemes that are perceptually assimilated to speech. The latter possibility would not necessarily preclude learning to *produce* clicks, at least in isolation. Non-click L1 learners of L2 click languages may be able to articulate the clicks without being able to perceive them as speech. However, our intuition is that if they do not perceive or produce clicks as speech, but only as nonspeech "mouth sounds," they will likely have difficulty coarticulating them properly with vowels, and hence producing them correctly in lexical items.

The L2 learning possibilities for non-assimilable phones suggest a particularly exciting line of future investigation on L2 perceptual learning. The extensions of PAM principles to the L2 learning situation thus offer a good range of empirical and theoretical issues for further investigation. The answers promise new insights into the perceptual side of L2 learning, complementing Flege's important contributions to understanding L2 perception and production within the SLM framework.

Concluding Comments: Considerations for Experiential Research on Perception

Among the key remaining questions for research and theory to address, regarding the relationship between nonnative speech perception in functional monolinguals and that in L2 learners is how variations in language input and the listener's level of L1 development (and L2, if being acquired) may influence perception of a given nonnative contrast. These considerations bring us back to our opening section, where we extended Heraclitus' river-crossing metaphor to the topic of experiential effects on speech perception. If both the language environment and the individual who is experiencing it are ever-changing, listener and input variations are likely to have substantive impact on perception of native and nonnative speech, especially if the perceiver is acquiring an L2.

When comparing naïve and L2 listeners, multiple factors are likely to differ between them beyond their difference in exposure to the target speech contrasts. With perception of a given nonnative contrast by naïve listeners versus by L2 learners of the target language, at the same age and thus the same number of years of native language (L1) development, one might be tempted to assume that the groups differ solely in exposure to the target nonnative/L2 contrast. However, they will certainly also differ on other dimensions: general effects of acquiring a non-native language that may (SLA) or may not (FLA) involve regular active contact with a non-native language environment and culture; exposure to native (SLA) and/or nonnative (FLA) pronunciation of the target and other speech elements; experience attempting to produce the target contrast; effects of establishing lexical items in the target language; and so on. In particular, establishing lexical items in the target language is likely to exert forceful linguistic pressure for the L2 learner to "re-phonologize" perception of the target contrasts, whereas the naïve listener has no such motivation. Relatedly, it would be quite informative to test whether some critical point in expansion of the L2 lexicon may indeed curtail perceptual learning, which appears most likely to take place relatively early in L2 acquisition, as we described earlier.

When evaluating changes in speech perception over learning or development, the passage of time is likely serving as a surrogate for the real underlying sources of influence: the listener's level of accomplishment in learning to recognize the linguistic structure and phonetic properties of their L1 and L2, and other individual and environmental factors that co-vary with their language development. The general developmental progression is that infants initially display universal capabilities to discriminate low-level properties of speech, then move on to recognizing certain language-specific phonetic details of the ambient language between 6-12 months, and then to an emerging recognition of contrastive phonologic aspects of lexical items by around 18-20 months. Further refinement in perceiving native phonetic and phonological properties occurs across the third and fourth years, with acquisition of reading skills at 5-6 years adding an orthographic-mapping contribution to greater phonological awareness (at least, for alphabetic orthographies, see e.g., Burnham, Tyler, & Horlyck, 2002). Approximately adult-like perception of both native and nonnative speech is apparent by about 8 years of age. On the input side, as noted earlier, conversational partners' speech displays marked adjustments in prosodic, phonetic, and syntactic properties according to the listener's developmental and linguistic characteristics. Ideally, future research should systematically examine both sources of influence on speech perception.

A number of other intriguing research questions are brought to the fore by the issues considered in this chapter. It will be of particular interest to investigate not only the PAM-L2 predictions we have offered, but also to fill various other gaps in the cross-language research literature. For example, as discussed earlier, it would be quite informative to examine perception of unfamiliar nonnative as well as native speech contrasts by L2 learners versus simultaneous versus early bilinguals; to examine L2 learners' and bilinguals' perception of L2 contrasts that constitute assimilation types other than Single Category assimilations in the L1; to compare perceptual attunement to L2 contrasts under classroom FLA versus immersion SLA contexts; to assess the posited narrow window for L2 speech perception learning in adults versus children, and to probe whether it interacts with L1 acquisition level in ultimate attainment of L2 speech perception. Additionally, key insights about fluent simultaneous or very early bilinguals would accrue from systematic examination of L1/dominant and L2/nondominant speech contrasts in language environments that differ in critical ways, for example, strongly bilingual environments, such as Montreal (French and English) and Barcelona (Spanish and Catalan), versus mixed multi-lingual environments, such as Vancouver, Canada (L1s of French, Cantonese, Hindi, etc.), and Sydney, Australia (Greek, Italian, Mandarin, Vietnamese, Arabic, etc.), where the majority of residents are monolingual speakers of a single dominant language but communities of numerous other L1s co-exist, for whom the dominant language is an early-learned L2.

We stand at an exciting threshold in research on experiential influences on speech perception. The coming decades of cross-language perception research promise to offer a wealth of new insights about L2 learners and multilinguals in various types of language environments.

References

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In B. Hála, M. Romportl & P. Janota (Eds.), *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569-573). Prague: Academia.

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics, 63*, 798-810.

Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). Point vowels in Japanese mothers' speech to infants and adults. *Journal of the Acoustical Society of America, 105*, 1095-1096.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*, 233-250.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 30*, 591-627.

Beddor, P. S., & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America, 71*, 1551-1561.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life.* (pp. 289-304). Dordrecht, The Netherlands: Kluwer Academic.

Best, C. T. (1994a). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The* *development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T. (1994b). Learning to perceive the sound pattern of English. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research* (Vol. 8, pp. 217-304). Hillsdale, NJ: Ablex Publishers.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Best, C. T., & Avery, R. A. (1999). Left hemisphere advantage for click consonants is determined by linguistic significance. *Psychological Science, 10*, 65-69.

Best, C. T., Faber, A., & Levitt, A. G. (1996). Perceptual assimilation of non-native vowel contrasts to the American English vowel system. *Journal of the Acoustical Society of America, 99*, 2602.

Best, C. T., Hallé, P. A., Bohn, O.-S., & Faber, A. (2003). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2889-2892). Barcelona: Causal Productions.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America, 109*, 775-794.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Best, C. T., Traill, A., Carter, A., Harrison, K. D., & Faber, A. (2003). !Xóõ click perception by English, Isizulu, and Sesotho listeners. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 853-856). Barcelona: Causal Productions.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11*, 303-328.

Bohn, O.-S., & Steinlen, A. K. (2003). Consonantal context affects cross-language perception of vowels. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2289-2292). Barcelona: Causal Productions.

Brancazio, L., Miller, J. L., & Paré, M. A. (2003). Visual influences on the internal structure of phonetic categories. *Perception & Psychophysics, 65,* 591-601.

Brannen, K. (2002). The role of perception in differential substitution. *The Canadian Journal of Linguistics/La Revue Canadienne de Linguistique, 47*, 1-46.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201-251.

Browman, C. P., & Goldstein, L. (1990). Representation and reality: physical systems and phonological structure. *Journal of Phonetics, 18*, 411-424.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155-180.

Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonology. In T. van Gelder & B. Port (Eds.), *Mind as motion* (pp. 175-193). Cambridge, MA: MIT Press.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research, 14*, 136-193.

Bundgaard-Nielsen, R. L., & Bohn, O.-S. (2004). Exploring the intelligibility of foreign-accented English vowels when “English” is ill-defined. *Journal of Acoustical Society of America, 115*, 2605.

Burnham, D., & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed.), *SouthEast Asian Linguistic Studies in Honour of Vichin Panupong* (pp. 29-47). Bangkok, Thailand: Chulalongkorn University Press.

Burnham, D., Kitamura, C., & Volmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296,* 1435.

Burnham, D., Tyler, M., & Horlyck, S. (2002). Periods of speech perception development and their vestiges in adulthood. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 281-300). Trier, Germany: Wissenschaftlicher Verlag Trier.

Burnham, D. K. (2000). Excavations in language development: Cross-linguistic studies of consonant and tone perception. In D. K. Burnham, S. Luksaneeyanawin, C. Davis & M. Lafourcade (Eds.), *Interdisciplinary approaches to language processing: The international conference on human and machine processing of language and speech*. Bangkok, Thailand: Chulalongkorn University Press.

Carroll, S. E. (1999). Putting "input" in its proper place. *Second Language Research, 15,* 337-388.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28,* 746-755.

Englund, K. T., & Behne, D. M. (2005). Infant directed speech in natural interaction--Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research, 34,* 259-280.

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

Flege, J. E. (1984). The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguistica, 18*, 117-138.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J. E. (1989). Chinese subjects’ perception of the word-final English /t/-/d/ contrast: Before and after training. *Journal of the Acoustical Society of America, 86*, 1684-1697.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-276). Timonium, MD: York Press.

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis.* (pp. 101-131). Mahwah, NJ: Lawrence Erlbaum Associates.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217-243). Trier, Germany: Wissenschaftlicher Verlag Trier.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Flege, J. E., & Hillenbrand, J. (1987). A differential effect of release bursts on the stop voicing judgments of native French and English listeners. *Journal of Phonetics, 15*, 203-208.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527-552.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*, 1-34.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*, 2973-2987.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*, 3623-3641.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*, 467-491.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99*, 1161-1173.

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97*, 2540-2551.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62*, 32-41.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “l” and “r”. *Neuropsychologia, 9*, 317-323.

Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech, 31*, 57-75.

Grieser, D. A., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*, 577-588.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*, 2711-2724.

Hallé, P. A., Best, C. T., & Bachrach, A. (2003). Perception of /dl/ and /tl/ clusters: A cross-linguistic perceptual study with French and Israeli listeners. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2893-2896). Barcelona: Causal Productions.

Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic versus phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics, 27*, 281-306.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*, 395-421.

Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). The processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance, 24*, 592-608.

Harnsberger, J. D. (2000). A cross-language study of the identification of non-native nasal consonants varying in place of articulation. *Acoustical Society of America, 108*, 764-783.

Harnsberger, J. D. (2001). On the relationship between identification and discrimination on non-native nasal consonants. *Acoustical Society of America, 110*, 489-503.

Ingram, J. C. L., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25*, 1-28.

Ingram, J. C. L., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English of /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America, 103*, 1-14.

Iverson, P. (2005). *Perceptual interference and learning in speech perception*. Paper presented at the ISCA Workshop on Plasticity in Speech Perception, London, UK.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553-562.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness of American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99*, 1130-1140.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47-B57.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics, 24*, 131-161.

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2005, accepted with revisions). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*.

Kahn, C. H. (1981). *Art and thought of Heraclitus*. Cambridge, UK: Cambridge University Press.

Kingston, J. (2003). Learning foreign vowels. *Language and Speech, 46*, 295-349.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy, 4,* 85-110.

Knoll, M., & Uther, M. (2004). Motherese and Chinese: Evidence for acoustic changes in speech directed at infants and foreigners. *Journal of the Acoustical Society of America, 116*, 2522.

Kochetov, A. (2004). Processing of secondary articulation contrasts in different syllable positions: Universal and language-particular factors. *Cahiers Linguistiques d'Ottawa, 32*, 24-43.

Kochetov, A., & So, C. K. (2005). *Perception of Russian stops by native speakers of Cantonese and English.* Paper presented at the First L2 Speech workshop of the Acoustical Society of America, Vancouver, Canada.

Kroll, J. F., & de Groot, A. M. B. (Eds.). (2005). *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Dordrecht: Kluwer Academic Publishers.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*, 684-686.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lindau, M. (1985). The story of /r/. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 157-168). New York: Academic Press.

Lisker, L., & Abramson, A. S. (1967). The voicing dimension: Some experiments in comparative phonetics. *Haskins Laboratories Status Report on Speech Research, SR11*, 9-15.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mother's speech clarity and infants' speech discrimination skills. *Developmental Science, 6*, F1-F10.

Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception & Performance, 23*, 1665-1679.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*, 271-285.

Miller, J. L., Connine, C. M., Schermer, T. M., & Kleunder, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, *73*, 2124-2133.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46*, 505-512.

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102*, 3742-3753.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of Acoustical Society of America, 89*, 2961-2977.

Polka, L. (1992). Characterizing the influence of native experience on adult speech perception. *Perception and Psychophysics, 52*, 37-52.

Polka, L. (1995). Lingustic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America, 97,* 1286-1296.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*, 221-231.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421-435.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy, 3*, 365-394.

So, C. K. (2005). The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *Journal of Acoustical Society of America, 117*, 2427.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America, 109*, 1691-1704.

Strange, W., & Bohn, O.-S. (1998). Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *Journal of the Acoustical Society of America, 104,* 488-504.

Sussman, J. E., & Gekas, B. (1997). Phonetic category structure of [I]: Extent, best exemplars, and organization. *Journal of Speech & Hearing Research, 40*, 1406-1424.

Sussman, J. E., & Lauckner-Morano, V. J. (1995). Further tests of the 'perceptual magnet effect' in the perception of (i): Identification and change-no-change discrimination. *Journal of the Acoustical Society of America, 97*, 539-552.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate nonnative speech sounds. *Canadian Journal of Psychology, 38*, 579-590.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. E. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics, 33,* 263-290.

Veleva, M. (1985). The perception of the voiced/voiceless phonological opposition in Bulgarian by Arab learners. *Supostavitelno Ezikoznanie/Contrastive Linguistics, 10*, 28-37.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, *92*, 723-735.

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, *44*, 95-118.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672-683.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percpetion & Psychophysics*, *37*, 35-44.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7,* 49-63.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, *25*, 501-528.

Yoneyama, K. (1997). A cross-linguistic study of diphthongs in spoken word processing in Japanese and English. *Ohio State University Working Papers in Linguistics, 50,* 163-175.

[1] Supported in part by grants from the U.S. National Institute on Deafness and Communication Disorders (NIDCD: DC00403) and the National Institute of Child Health and Development (NICHD: HD01994). While the authors take sole responsibility for what is written on these pages, many thanks are due to colleagues and students for their support, assistance, and inspiration throughout the work reported here, including Louis Goldstein, Carol Fowler, Doug Whalen, Alice Faber, Michael Studdert-Kennedy, and Al Liberman. Thanks are due as well to many students and research assistants for their inspiration and hard work, including Nomathemba Sithole Shepherd, Robert Avery, Janet Calderón, and Tiffany Gooding.

[2] Authors' address: MARCS Auditory Laboratories, U. Western Sydney, Penrith South DC NSW 1797 Australia. e-mails: c.best@uws.edu.au and m.tyler@uws.edu.au. The first author is also a Senior Scientist at Haskins Laboratories, 300 George Street, New Haven CT 06511 U.S.A. e-mail: best@haskins.yale.edu

[3] This does *not* rule out relatively passive exposure to a language other than the L1, that is, for which the listener has made little or no active attempt to learn the language. Nor does it rule out limited L2 instruction, especially classroom-only instruction with instructors who have a strong L1 accent.

[4] Note, however, that the pattern of asymmetry and perceptual warping has failed replicaiton in several laboratories, and the phenomenon and/or its explanation have been questioned by some researchers (Lively & Pisoni, 1997; Sussman & Gekas, 1997; Sussman & Lauckner-Morano, 1995).

[5] There also has been variability in definition of "late onset," though generally it is assumed to be after puberty, for example, after 13 years (or 15, or 19-20). As age of onset has been the subject of numerous discussions in the literature, we will instead focus on the issue of how much SL experience counts as "experienced."

[6] We claim this only for *perceptual* modifications, not for changes in production, which may continue to adjust to L2 phonetic and phonological properties over a longer time span (though rarely to native-like levels), as work by Flege and others has indicated. The implied difference in L2 learning on the perceptual versus the productive side could converge toward support of claims that "perception precedes production" and that "perceptual abilities underlie production abilities in L2", but a definitive answer requires more systematic examination of LOR (length of residence) effects on perception.

[7] However, the time frame for learning may vary according to type of contrast (e.g., stop voicing vs. place of articulation: Tees & Werker, 1984).

---

[8] Two other assimilation patterns for nonnative contrasts are also possible but have not been described in detail in treatments of PAM, for example, Uncategorized-Nonassimilable, Categorized-Nonassimilable, either of which should be discriminated well as speech versus nonspeech distinctions.