15

The phonetics of second language learning and bilingualism

Charles B. Chang

Introduction

How are speech sounds and patterns mastered in a second language (L2), especially when the L2 is learned later in life? This question is at the heart of research in L2 speech learning, an interdisciplinary field at the nexus of phonetics, phonology, cognitive psychology, L2 acquisition, and applied linguistics. The broad goal of L2 speech research is to understand the mechanisms and processes underlying L2 speech development, with a view toward applications in language learning and language pedagogy. This chapter provides an overview of the major theories and findings in the field of L2 speech learning. For reasons of space, the discussion focuses primarily on four main conceptual frameworks, among the most detailed and widely tested in the field: the Perceptual Assimilation Model – L2, the Native Language Magnet Theory, the Automatic Selective Perception Model, and the Speech Learning Model. These frameworks differ in terms of empirical focus, including the type of learner (e.g., beginner vs. advanced) and target modality (e.g., perception vs. production), and in terms of theoretical assumptions, such as the basic unit or window of analysis that is relevant (e.g., articulatory gestures, position-specific allophones).

To evaluate the predictive differences among these theories, this chapter discusses a number of empirical studies that have investigated L2 speech primarily at a segmental level. However, it should be pointed out that research on L2 speech learning addresses many different aspects of speech, including overall accent (e.g., Yeni-Komshian, Flege, and Liu, 2000), segment sequences (i.e., phonotactics; Dupoux, Hirose, Kakehi, Pallier, and Mehler, 1999; Altenberg, 2005a; Davidson, 2006; Hallé and Best, 2007), and higher-level prosodic structure (e.g., word boundaries and cross-word sandhi phenomena; Altenberg, 2005b; Zsiga, 2011; Schwartz, 2016). Further, because other recent publications in L2 phonetics and phonology already provide extensive reviews of empirical findings in this area (e.g., Eckman, 2012; Broselow and Kang, 2013; Colantoni, Steele, and Escudero, 2015; Simonet, 2016; Davidson, 2017; Bohn, 2017), the current contribution is oriented instead toward presenting a synthesis of theoretical approaches to the study of L2 speech. Thus, although several empirical studies are covered in a fair amount of detail, we will concentrate primarily on exploring the points of convergence and divergence, as well as the complementarities, among theories of L2 speech.

Despite the ways in which theories of L2 speech differ from one another, three recurring themes emerge from the L2 speech literature. First, the learning of a target L2 structure (segment, feature, phonotactic constraint, prosodic pattern, etc.) is influenced by phonetic and/or phonological similarity to structures in the native language (L1). In particular, L1-L2 similarity exists at multiple levels and does not necessarily benefit L2 outcomes. Second, the role played by certain factors, such as acoustic-phonetic similarity between close L1 and L2 sounds, changes over the course of learning, such that advanced learners may differ from novice learners with respect to the effect of a given variable on observed L2 behavior. Third, the connection between L2 perception and production (insofar as the two are hypothesized to be linked) differs significantly from the perception-production links observed in L1 acquisition. Each of these themes is addressed in more detail in the rest of the chapter.

As an interdisciplinary area of inquiry, L2 speech research is intrinsically linked not only to experimental advances in phonetics and laboratory phonology, but also to theoretical views of the process and outcomes of L2 acquisition. Crucially, the view of L2 acquisition adopted in this chapter is one that identifies the start of L2 acquisition with the onset of bilingualism. That is, the chapter considers L2 learners both as acquirers of a new language and as individuals with two languages, for two reasons. First, many individuals exposed to an L2 will eventually become proficient bilinguals, and there is no clear dividing line between "L2 learner" and "bilingual." Second, given that the majority of the world can be described as bilingual or multilingual (Tucker, 2001), bilingualism, rather than monolingualism, may be the most appropriate point of departure for considering interlanguage phenomena in L2 learning. As such, this chapter situates the study of L2 speech within the long tradition of bilingualism research on bidirectional cross-linguistic interactions. Under this view, it benefits L2 speech research to consider not only L2 outcomes, but also the manner in which learners' developing L2 knowledge may influence their knowledge and use of the L1 (e.g., Cook, 2003). Thus, it should be noted that, although not covered here in detail, the burgeoning literature on L1 phonetic and phonological change in L2 learners (for reviews, see Celata, in press; Chang, in press; de Leeuw, in press) is relevant to the study of L2 speech because it can provide unique insights into learners' observed trajectory of L2 development.

In the rest of this chapter, we review the principles of the four selected conceptual frameworks for the study of L2 speech and discuss two topics that remain areas of active research in L2 phonetic learning and bilingualism: (a) the role of the L1 – in particular, the type and degree of similarity between the L1 and L2 – in L2 phonetic development, and (b) links between L2 perception and L2 production.

Theoretical frameworks

The theoretical review first examines frameworks focusing on L2 perception, and then proceeds to the main framework addressing L2 production. We begin with arguably the most widely tested theory of nonnative and L2 speech perception, the Perceptual Assimilation Model – L2.

The Perceptual Assimilation Model – L2

One of the most influential theories of L2 speech perception, the Perceptual Assimilation Model – L2 (PAM-L2) was proposed by Best and Tyler (2007), expanding upon an earlier theory of nonnative speech perception focused on naive listeners, the Perceptual

Table 15.1 Types of perceptual (non)assimilation posited in PAM-L2. TC: Two Category, CG: Category Goodness, SC: Single Category, UC: Uncategorized-Categorized, UU: Uncategorized-Uncategorized, NA: Non-Assimilable. Accuracy scale: 1 = poor, 2 = fair, 3 = good, 4 = excellent.

	TC	*CG*	*SC*	*UC*	*UU*	*NA*
L2 sounds perceived as speech?	yes	yes	yes	yes	yes	no
L2 sounds both assimilated to L1?	yes	yes	yes	no	no	no
Difference in goodness-of-fit to L1?	no	yes	no	yes	–	–
Discrimination accuracy (1–4 scale)	4	2–3	1	4	1–2	3–4

Assimilation Model (PAM; Best, 1994). PAM-L2 expands upon PAM by incorporating a role for the additional knowledge that L2 learners, but not naive monolinguals, have about the target language (e.g., phonological knowledge). However, the two models are similar in their assumption of articulatory gestures as the basic phonetic unit (a view following from the direct realist approach to speech perception; for further discussion, see Best, 1995, pp. 172–175), in their focus on perception, and in their account of L2 perceptual patterns in terms of *perceptual assimilations* to L1 sounds.

The core logic of PAM-L2 is that sounds of an L2 are differentially discriminable according to how they are perceptually assimilated to (i.e., interpreted in terms of) L1 sounds. The theory sets out a typology of diverse ways in which two L2 sounds *x* and *y* can be assimilated to L1 sounds, which lead to different degrees of success discriminating *x* and *y* (see Table 15.1). In the case of Two Category (TC) assimilation, *x* and *y* are assimilated to two different L1 sound categories and, given that there is no pressure from the L1 toward perceptual conflation of *x* and *y*, L2 learners are able to discriminate them with a high degree of accuracy. On the other hand, in Single Category (SC) assimilation, *x* and *y* are assimilated to the same L1 sound category, each with a similarly high goodness-of-fit (i.e., *x* and *y* are both phonetically close to the L1 attractor); in this case, there is strong pressure from the L1 toward perceptual conflation of *x* and *y*, and consequently L2 learners discriminate them poorly. Finally, in Category Goodness (CG) assimilation, *x* and *y* are assimilated to the same L1 sound category, but with different degrees of fit (i.e., *x* and *y* are unequally close to the L1 attractor), leading to discrimination performance that is intermediate between the TC and SC cases (ranging from fair to good). For example, for L1 English listeners, the Zulu contrasts between lateral fricatives /ɬ/ and /ɮ/, bilabial stops /b/ and /ɓ/, and velar stops /kʰ/ and /k'/ were predicted, respectively, to undergo TC assimilation (to a voiceless English fricative such as /ʃ/ and a voiced English fricative such as /ʒ/), CG assimilation (both to English /b/, with different degrees of fit), and SC assimilation (both to English /k/, with similar degrees of fit), and indeed English listeners' discrimination of these contrasts was, respectively, good, somewhat less good, and poor (Best, McRoberts, and Goodell, 2001).

In addition to the TC, SC, and CG types of perceptual assimilation, which all involve both members of the L2 sound contrast being assimilated to L1 sounds, it is possible for one or both members of an L2 contrast not to be assimilated to L1 sounds. In Uncategorized-Categorized (UC) assimilation, only one of the L2 sounds is identified with an L1 sound, while the other is perceived as unlike any L1 sound. Because here the L2 contrast reduces to an opposition between L1-like and non-L1-like sounds, L2 learners are able to discriminate the sounds well (e.g., Thai back unrounded vowels /ɯ/ and /ɤ/ for L1 English listeners; Tyler, Best, Faber, and Levitt, 2014). However, in Uncategorized-Uncategorized

(UU) assimilation, neither of the L2 sounds is perceived as similar to an L1 sound. In this case, therefore, there is no possibility of using a framework of "L1-like" vs. "non-L1-like," and discrimination accuracy ranges from poor to intermediate, depending on the degree to which the members of the contrast (partially) resemble the same L1 sound or different L1 sounds.

Finally, it is possible for L2 sounds not only to resist identification with an L1 sound, but also to be perceived as non-linguistic. In the rare Non-Assimilable (NA) case of L2 sounds, the L2 sounds are so divergent from any member of the L1 sound inventory that they are effectively treated as non-speech. In this case, discrimination of the L2 contrast may benefit from a non-speech mode of auditory processing that has not been warped by linguistic categories (see next section, *The Native Language Magnet Theory*). This possibility was elegantly demonstrated in a series of studies on nonnative click perception (Best, McRoberts, and Sithole, 1988; Best, Traill, Carter, Harrison, and Faber, 2003). Contrary to the hypothesis that nonnative clicks would be most easily discriminated by speakers already familiar with clicks from their L1, speakers with L1 clicks tended to do relatively poorly on nonnative click discrimination; by contrast, L1 English speakers who had no experience with clicks performed relatively well. The explanation for this result lies in the fact that click-language speakers are subject, at least some of the time, to SC assimilation due to L1 click categories that serve as perceptual attractors for nonnative clicks. Non-click-language speakers such as L1 English speakers, however, are unencumbered by L1 click categories and, furthermore, have nothing in their L1 inventory that remotely resembles clicks auditorily. Therefore, they are free to discriminate the nonnative click contrasts purely in terms of their acoustic characteristics.

As for the basis of perceptual assimilation of L2 sounds to L1 sounds, PAM-L2 posits that L2-to-L1 mapping may occur due to cross-linguistic similarity at a gestural level as well as at a phonological level. Indeed, for L2 learners (as opposed to non-learners) abstract phonological knowledge of the L2 is likely to play an important role in establishing equivalences between L1 and L2 sounds; we will return to this topic later in the chapter. Broadly speaking, however, the basis for perceptual assimilation can be viewed as proximity between L2 sounds and L1 attractors. In the next section, we introduce a theory which, like PAM-L2, formalizes the role of L1 attractors in speech perception, but with respect to the internal structure of L1 phonological categories and the acoustic perceptual space.

The Native Language Magnet Theory

Developed originally to account for L1 perceptual development in infants, Kuhl's Native Language Magnet (NLM) Theory has been applied to the study of L2 speech as well (Kuhl and Iverson, 1995; Ellis, 2006). In particular, the revised and expanded Native Language Magnet Theory (NLM-e) enumerates a number of basic principles, several of which help account for aspects of L2 perception (Kuhl et al., 2008). First, the L1 learner progresses from perceiving speech in a universal manner (i.e., not specialized for the L1) to perceiving it in an L1-specific manner, a transition that is driven both by the detection of distributional patterns in the ambient L1 input and by the enhanced acoustic properties of infant-directed speech. Second, exposure to the L1 leads to a *neural commitment* to L1 speech patterns, which biases future speech learning (cf. the "selective perception routine" of the Automatic Selective Perception Model, discussed in the next section). Third, L1 phonetic learning is influenced by social interaction. Fourth, L1 acquisition involves forming links between speech perception and speech production (i.e., perception-production links are developed through experience rather than being innate).[1] Fifth, early perceptual abilities for, as well as neural responses to, native and nonnative sound contrasts are predictors of L1 development. In particular, better perceptual abilities (e.g., better discrimination performance) for the L1 predict faster L1 development, whereas better perceptual abilities for nonnative speech predict slower L1 development.

The first and second of the preceding principles relate to two core concepts relevant for the study of L2 speech: *perceptual warping* and the *perceptual magnet*. Perceptual warping refers to the way in which the acoustic perceptual space related to a given type of speech sound (e.g., a multidimensional formant space in the case of vowels) is transformed with the accumulation of linguistic (i.e., L1) experience, while a perceptual magnet is a specific part of that modified perceptual space – namely, the prototype of a contrastive sound category. L1 learners develop such prototypes for speech sounds early in life during their distributional analysis of L1 input, and these prototypes act as attractors for newly perceived speech tokens, leading to the observation of a so-called "perceptual magnet effect" in humans only (Kuhl, 1991). The perceptual magnet effect describes essentially the same type of phenomenon as perceptual assimilation (as in PAM-L2), except that the perceptual magnet effect does not refer to cross-linguistic assimilation per se; rather, the idea is that, once there are phonological categories in place, listeners are biased to perceive incoming speech input in terms of these categories as opposed to objectively, without reference to categories. This effect thus underlies both the "categorical perception" of L1 speech (Liberman, Harris, Hoffman, and Griffith, 1957) as well as the perceptual assimilation of L2 sounds to L1 categories.

Crucially, however, the strength of the perceptual magnet effect differs according to proximity to a category prototype (i.e., the magnet). That is to say, a speech token, regardless of its source language, is more likely to be perceived in terms of an L1 category the closer it is to the category prototype. This pattern relates back to the notion of perceptual warping: with L1 experience and the development of prototypes, the acoustic perceptual space becomes "warped," with the result that a given phonetic distance is perceived as smaller when close to a prototype than when far from a prototype. The reason for this phenomenon is the nature of a prototype's "gravitational pull," which diminishes in strength as one moves further away from the prototype.

Applied to L2 speech perception, the perceptual warping involved in L1 development provides a converging, yet slightly different, account for many of the same findings as PAM-L2, such as the lower discriminability of SC contrasts compared to CG contrasts (recall from previous discussion the Zulu contrasts /k^h/-/k'/ and /b/-/ɓ/, respectively SC and CG for L1 English listeners). In the case of an SC contrast and a CG contrast whose members are equally far apart phonetically, the SC contrast, by virtue of the fact that both members are by definition very close to the L1 category to which they are assimilated, will necessarily be closer overall to the L1 category prototype than the CG contrast. Therefore, the phonetic distance represented in the SC contrast will be harder to perceive than the (equal) phonetic distance represented in the CG contrast, because the former is more strongly pulled into the L1 category.

In short, the NLM(-e) view of L1 development in terms of prototype formation and perceptual warping formalizes crucial outcomes of L1 experience that have consequences for L2 perception. In the next section, we review a theory of L2 perception which is similar to NLM in terms of formalizing outcomes of L1 experience and additionally draws an explicit link to L2 perception.

The Automatic Selective Perception Model

Like NLM, the Automatic Selective Perception (ASP) Model of L2 speech perception (Strange, 2011) understands L1 biases in L2 perception as the outcome of a process of perceptual specialization for the L1. According to ASP, perceptual specialization involves establishing *selective perception routines* (SPRs) that allow perception to be targeted, automatic, and robust in adverse conditions. Consistent with the NLM view of perceptual specialization for the L1 as a central component of L1 perceptual ability, ASP views the development of L1-appropriate SPRs as critical to becoming a skilled L1 listener. However, these L1 SPRs also lead to L1 interference in perception of an L2, because the L2 will often require learners to attend to different properties of the speech signal than the ones relevant in their L1, and/or to integrate these cues differently.

Crucially, in ASP (consistent with the SLM and theories of L1 phonological development such as the Processing Rich Information from Multidimensional Interactive Representations framework; Werker and Curtin, 2005), the unspecialized (i.e., language-general) processing abilities evident in childhood remain available throughout life (and, therefore, in adulthood as well). However, when the cognitive demands of a language task are high (e.g., in processing a syntactically complex L2 utterance), L2 learners' access to these abilities may be blocked, resulting in a default to automatized L1 SPRs. Thus, the effect of task demands on processing L2 speech is a core consideration of ASP, which distinguishes this theory from NLM.

Another aspect of ASP that distinguishes it from NLM is the explicit link it draws between stage of L2 acquisition and manner of L2 perceptual processing. ASP identifies L2 experience as a factor influencing L2 perception, consistent with studies suggesting that advanced L2 learners tend to perceive the L2 significantly differently from novice L2 learners – namely, in a more "phonologized" manner. This process of phonologization was apparent, for example, in Levy and Strange's (2008) study of experienced and inexperienced L2 listeners of French, both from an L1 English background. These L2 listeners were tested on discrimination of several French vowel contrasts, including front rounded vowels that do not occur in English (/y/, /œ/) and vowels occurring in both bilabial and alveolar contexts. Results showed two systematic disparities between the experienced and inexperienced groups. First, with the exception of /u/-/y/, the experienced listeners outperformed the inexperienced listeners overall on most of the vowel contrasts. Second, there was a significant context effect for inexperienced listeners, but not for experienced listeners: inexperienced listeners performed differently on certain vowel contrasts across coarticulatory contexts (e.g., higher error rate on /u/-/y/ in an alveolar context than in a bilabial context). These findings are consistent with the view that L2 speech learning involves developing distinct representations for new L2 sounds (e.g., /y/ and /œ/ for L1 English speakers) as well as familiarity with rule-governed coarticulatory patterns in the L2, which allows learners to abstract over phonemically non-contrastive coarticulatory variation, such as the vowel fronting effect associated with alveolars.

Findings in studies such as Levy and Strange (2008) converge with the results of many other studies (e.g., Bradlow, Pisoni, Akahane-Yamada, and Tohkura, 1997; Wang, Spence, Jongman, and Sereno, 1999; Aoyama, Flege, Guion, Akahane-Yamada, and Yamada, 2004; Tajima, Kato, Rothwell, Akahane-Yamada, and Munhall, 2008) in suggesting that L2 experience generally helps the L2 learner to become more skilled at perceiving the L2 (though see Holliday, 2016 for an interesting counterexample from L1 Mandarin learners of Korean). This positive correlation between L2 experience and L2 perceptual performance can be attributed to two beneficial, and related, outcomes of L2 speech learning: (a) development of mental representations for the contrastive sounds of the L2, particularly those which do not occur in the L1, and (b) development of SPRs for the L2. Both of these developments allow L2 perception to be more targeted, automatic, and robust, resulting in a significant advantage for experienced L2 listeners compared to naive or inexperienced listeners.

Thus, ASP accounts for L1 biases in L2 perception, as well as for L2 perceptual learning over time, in terms of the same fundamental construct: SPRs, which direct a listener's attention to a proper subset of the numerous acoustic properties that a listener could potentially attend to in the speech signal. ASP differs from PAM-L2 and NLM in focusing more on cue weighting than on cross-linguistic mapping or category prototypes; this focus helps to account for perceptual variation observed among L2 learners with similar L1 phonological constraints (see, e.g., Chang, 2018). However, all three of these theories are similar in that they are theories of L2 perception, not theories of L2 production. Next, we discuss a theory of L2 speech that addresses aspects of both perception and production.

The Speech Learning Model

Unlike the theories discussed above, Flege's (1995, 1996, 2002) Speech Learning Model (SLM) is a theory of both L2 perception and L2 production. Its account of L2 speech consists of six main tenets, the first three being that (a) language learners maintain continuous access to the same basic learning mechanisms over the lifespan (i.e., adult learners are not fundamentally different from child learners in this respect), (b) L1 and L2 sounds exist in a "common phonological space" (Flege, 1995, p. 239), and learners are generally motivated to maintain cross-linguistic contrast between them, and (c) there is a tendency for *equivalence classification* of L2 sounds with close L1 counterparts. This mechanism of equivalence classification is not specific to L2 learning, but rather is used in the L1 to abstract appropriately over phonetic variability in L1 speech. The inappropriate operation of equivalence classification in L2 speech learning, however, may result in problems with perception and/or production of target L2 sounds.

A fourth, and central, claim of the SLM is that L2 sounds are differentially difficult to learn, depending on their phonetic proximity to L1 sounds (see Figure 15.1). In particular, the SLM posits three types of L2 sounds – *identical*, *new*, and *similar* – which form a hierarchy of learning difficulty as follows (from least to most difficult): identical < new < similar. Identical sounds are the least difficult to learn because in all relevant aspects they are exactly the same as their closest L1 sound; therefore, straight transfer of the L1 sound to the L2 will result in high accuracy with the L2 sound immediately. New sounds, by contrast, are more difficult to learn because, although they resist equivalence classification with L1 sounds due to a high degree of disparity along one or more dimensions, this cross-linguistic disparity also requires some novel aspects of the L2 sound to be learned to approximate target-like performance. These novel aspects, however, are hypothesized to be learnable in the long term. On the other hand, similar sounds are the most difficult to learn because they are close enough to L1 sounds to undergo equivalence classification with them, yet far enough from L1 sounds that simple transfer of the L1 sounds is not sufficient for target-like performance. In other words, similar sounds exist in an intermediate space of cross-linguistic similarity (as shown in Figure 15.1), which introduces the possibility of inappropriate influence from properties of close L1 sounds.

The nature of L1 influence for the three types of L2 sounds is captured in the fifth tenet of the SLM: L2 sounds may either approximate (i.e., assimilate properties of) or dissimilate

Figure 15.1 Continuum of similarity of L2 sounds to L1 sounds. NEW sounds are the least similar to L1 sounds; IDENTICAL sounds, the most similar; and SIMILAR sounds, intermediate in similarity.

from L1 sounds. In particular, when an L2 sound undergoes equivalence classification with a close L1 sound (as in the case of identical and similar sounds), the L1 and L2 sounds become "diaphones," sounds that are *perceptually linked* in the mind of the L2 learner. In the case of a similar L2 sound, this perceptual linkage (i.e., partial or total co-representation of the L2 sound with the L1 sound) eventually leads to the L1 and L2 sounds approximating each other in production (e.g., Williams, 1979; Major, 1992). On the other hand, when an L2 sound avoids equivalence classification with L1 sounds (as in the case of new sounds), the L2 sound is represented distinctly from L1 sounds. This distinct representation allows the L2 sound to be produced eventually in a target-like manner (e.g., Williams, 1977; Antoniou, Best, Tyler, and Kroos, 2010). Alternatively, however, L2 sounds represented distinctly from the closest L1 categories may dissimilate from them so as to maximize cross-linguistic contrast within the shared L1-L2 phonological space; such dissimilation may also result in L2 sounds diverging from native (monolingual) norms. For example, whereas early L1 Korean-L2 English bilinguals were found to produce native-like voice onset time (VOT) in their L2 voiceless stops (Kang and Guion, 2006), an early L1 French-L2 English bilingual produced L2 VOTs that were longer than native (i.e., past the English monolingual norm), in an apparent attempt to differentiate the L2 stops from the L1 stops (Mack, 1990).

Although assimilation and dissimilation result in opposite directions of movement relative to an L1 sound, crucially they may affect sounds of both the L1 and the L2, in line with the sixth tenet of the SLM: cross-linguistic influence (CLI) is, in principle, bidirectional. Thus, CLI is not limited to L1 sounds influencing L2 sounds (i.e., "forward transfer"), but may also result in L2 sounds influencing L1 sounds (i.e., "backward transfer"). Bidirectionality of CLI at the phonetic level was shown in a seminal study of English-French bilinguals, L1 English late learners of French and L1 French late learners of English (Flege, 1987). This study focused on two acoustic properties of learners' speech in both the L1 and the L2: the VOT of /t/ (canonically short-lag in French, but long-lag in English) and the second formant (F_2) frequency of /u/ (canonically low in French, but high in English) as well as /y/, a vowel phoneme that occurs only in French. Results provided evidence for bidirectional CLI. On the one hand, many (but not all) L1 English learners of French produced French /t/ with too-long VOTs (i.e., as English-influenced) and English /t/ with too-short VOTs (i.e., as French-influenced); a similar pattern was found with the VOTs of L1 French learners of English. As for F_2, both L1 English learners of French and L1 French learners of English produced French /u/ with too-high F_2 values (i.e., as English-influenced), although only the L1 French learners of English also produced English /u/ with too-low F_2 values. Notably, the L1 English learners additionally managed to produce the French /y/ with native-like F_2 values. Thus, overall, the pattern of results in this study was consistent with the SLM in two main respects: (a) showing a disparity between "similar" sounds (e.g., English and French /t/), which evince CLI due to equivalence classification, and "new" sounds (e.g., French /y/ for L1 English learners), which avoid CLI, and (b) showing bidirectional CLI.

Two aspects of the SLM that distinguish it from other influential theoretical frameworks for L2 speech research are its explicit application to production (as opposed to a focus on perception) and its prediction of bidirectional CLI (as opposed to unidirectional CLI, specifically L1 influence on the L2). These aspects of the framework make it especially appropriate for studies of L2 production (as in Flege, 1987) and studies of L1 change in L2 learners, both in production (e.g., Major, 1992, 1996; Chang, 2011, 2012b, 2013; de Leeuw, Schmid, and Mennen, 2010; Dmitrieva, Jongman, and Sereno, 2010) and perception (e.g., Tice and Woodley, 2012; Ahn, Chang, DeKeyser, and Lee-Ellis, 2017).

Summary and synthesis

In short, while the theories discussed at the beginning of this chapter often make convergent predictions in regard to L2 phonetic development, they differ in a number of ways. The primary dimensions of difference among these theoretical frameworks are summarized in Table 15.2, including the L2 experience or proficiency level of the learner described by the theory, the basic unit of analysis, the L2 domain(s) covered, and the foundation of the theory's explanation of CLI in either direction (L1-to-L2, L2-to-L1).

As discussed above, CLI at the phonetic level has been of special concern in the study of L2 speech and bilingualism, spawning a wealth of findings on L1-L2 phonetic interaction in L2 learners of various backgrounds (for further reviews, see Mack, 2003 and Kartushina, Frauenfelder, and Golestani, 2016). Apart from showing L1 influence in their L2 production (e.g., Port and Mitleb, 1983; Gass, 1984), learners may produce neither the L1 nor the L2 as native-like (Flege and Eefting, 1987b); they may also show little to no phonetic differentiation between the two languages (e.g., Williams, 1979; Major, 1992), leading to "compromise" values between L1 and L2 norms (e.g., Flege, 1987). On the other hand, learners' production of the L2, as well as of the L1, may also be native-like (e.g., Fokes, Bond, and Steinberg, 1985; Mack, 1989), and especially for early bilinguals, close approximation of monolingual norms for both languages is clearly possible (e.g., Caramazza, Yeni-Komshian, Zurif, and Carbone, 1973; Williams, 1977; Flege and Eefting, 1987a; Kang and Guion, 2006; Antoniou et al., 2010). All of the theories discussed in this chapter address CLI in the

Table 15.2 Comparison of selected frameworks for L2 speech research. PAM-L2: Perceptual Assimilation Model-L2, SLM: Speech Learning Model, NLM: Native Language Magnet Theory, ASP: Automatic Selective Perception Model. CLI: cross-linguistic influence. NA: not applicable.

	PAM-L2	*NLM*	*ASP*	*SLM*
Learner level	novice to advanced	advanced	novice to advanced	advanced
Basic unit	articulatory gesture	phonological category	auditory cue	position-specific allophone
About perception?	yes	yes	yes	yes
About production?	no	no	no	yes
Account of L1 → L2 CLI?	perceptual assimilation	perceptual warping	perceptual attunement	L1-L2 diaphones
Account of L1 ← L2 CLI?	NA	NA	NA	L1-L2 diaphones

L1-to-L2 direction, but not necessarily in the L2-to-L1 direction. In this regard, the SLM is unique in providing an account of bidirectional CLI.

To close this section, it is worth noting that the diversity of L1 and L2 outcomes in L2 learners has been approached analytically in additional ways, including systems typology and computational modeling. In regard to typology, Laeufer (1996) presents an attempt to schematize the different possible bilingual phonological systems, which each lead to a specific pattern of L1 and L2 speech production. Combining aspects of the bilingual lexical/conceptual model of Weinreich (1953) with the tripartite speech production model of Keating (1984), Laeufer distinguishes among three types of bilingual phonological system (coexistent, merged, and super-subordinate) in terms of the conflation of the L1 and L2 at various levels of representation (for further discussion, see Chang, 2010a, pp. 49–54). In regard to modeling, Tobin, Nam and Fowler (2017) provide an example of a formal computational account of shifts in bilingual speech through variation in the ambient language environment. This computational account assumes a dynamical systems framework, increasingly common in research on language development and change (see, e.g., de Bot, Lowie, and Verspoor, 2007; de Leeuw, Mennen, and Scobbie, 2013), and is also consistent with exemplar approaches to phonology and L2 acquisition incorporating a role for episodic memory (Johnson, 1997; Pierrehumbert, 2001; Hazan, 2007).

The role of L1-L2 similarity

Although the theories of L2 speech discussed in this chapter differ in a number of ways, what they have in common is the acknowledgment and incorporation of a role for the L1 in L2 development. Whether described as perceptual assimilation to the L1, equivalence classification with the L1, or simply transfer of an L1 category space or L1 selective perception routines, aspects of the L1 are taken to exert a powerful influence on L2 speech. The SLM and PAM-L2 in particular are based on cross-linguistic mapping of L2 sounds to L1 sounds, which raises the question of how L2 learners identify the L1 correspondents of L2 sounds. In other words, assuming that the main criterion for establishing L1-L2 correspondence is linguistic similarity, how do L2 learners make judgments of similarity between L1 and L2 sounds?

At the heart of this question is a crucial feature of L2 learners that distinguishes them from naive listeners: abstract knowledge of the target language. As acknowledged in PAM-L2, unlike naive listeners (who, by definition, are not familiar with the L2), L2 learners may have a considerable amount of higher-level knowledge of the L2, including knowledge of the phonemic inventory, phonotactic constraints, allophonic alternations, and/or the orthographic system used to visually represent the sounds of the language (cf. Polka, 1991, 1992; Best and Tyler, 2007; Boomershine, Hall, Hume, and Johnson, 2008; Davidson, 2011). Consequently, there are several sources of information about L2 sounds that L2 learners may take into account in forming a judgment of L1-L2 similarity that go beyond the raw phonetic data available to both L2 learners and naive listeners (which is generally taken to be gestural in PAM-L2 and acoustic in the SLM).

For L2 learners, the availability of multiple sources of information about L2 sounds introduces the possibility of conflict between those sources (Chang, 2015). To take one example pointed out by Chang, Yao, Haynes, and Rhodes (2011), which parallels the situation in Flege (1987), high rounded vowels in Mandarin Chinese (i.e., /y/ and /u/) resemble American English vowels at two levels: (a) an acoustic level (e.g., F_2 frequency, formant trajectories), and (b) a phonemic level (e.g., being a high back rounded vowel, /u/). In terms of acoustic proximity in $F_1 \times F_2$ space, Mandarin /y/ is closer to English /u/ than is Mandarin /u/, due to the fact that (American) English /u/ tends to be relatively front and, in certain dialects such as Southern Californian English, phonetically unrounded as well (Hagiwara, 1997; Ladefoged, 1999).[2] However, in terms of phonemic correspondence, Mandarin /u/ is the closer match to English /u/, because Mandarin /u/ (and not /y/) is a high back rounded vowel like English /u/; this is reflected, for example, in similar phonotactic patterning in the two languages (e.g., nonoccurrence of onset C/wu/).

Thus, in the case of L1 English learners of Mandarin, production of the L2 Mandarin vowels /y/ and /u/ could be influenced by the properties of English /u/ in at least two different ways, depending on whether acoustic or phonemic considerations take precedence in cross-linguistic mapping. First, Mandarin /y/ could be mapped to English /u/, with Mandarin /u/ avoiding equivalence classification with an English vowel; this would lead to too-low F_2 values for Mandarin /y/ (as English /u/ is characterized by a lower F_2 than Mandarin /y/) and target-like production of Mandarin /u/ (a "new" sound in SLM terms). Alternatively, Mandarin /u/ could be mapped to English /u/, with Mandarin /y/ avoiding equivalence classification with an English vowel; this would lead to too-high F_2 values for Mandarin /u/ (as English /u/ is characterized by a higher F_2 than Mandarin /u/) and a target-like production of Mandarin /y/. What Chang et al. (2011) found was the latter: relatively experienced L1 English learners of Mandarin produced Mandarin /u/ with F_2 values that were higher than those of L1 Mandarin speakers, but Mandarin /y/ with F_2 values not significantly different from those of L1 Mandarin speakers, replicating the pattern reported in Flege (1987) for advanced L1 English learners of French.

Together with a range of other results from L2 perception (Polka and Bohn, 1996; Strange, Levy, and Lehnholf, 2004), loanword adaptation (Kang, 2008; Chang, 2009, 2012a), and phonetic drift (Chang, 2012b, 2013), findings such as in Flege (1987) and Chang et al. (2011) contribute to a picture in which cross-linguistic mapping (i.e., relating an L2 sound to an L1 sound) often follows the phonemic route over the acoustic route. To take one other example from loanword adaptation, high lax vowels in English loanwords are adapted by French-English bilinguals not with the acoustically closest mid vowels of French, but with the phonemically parallel (and acoustically more distant) high vowels of French (LaCharité and Paradis, 2005). The fact that this happens with both of the English vowels at issue (/ɪ/ and /ʊ/), not just with one or the other, suggests that cross-linguistic mapping on a phonemic basis is systematic rather than idiosyncratic. Thus, at least for advanced L2 learners, there may be a privileged status of abstract (higher-level) information about L1 and L2 sounds in making judgments of cross-linguistic similarity. However, the specific dynamics of interaction between phonetic and phonological types of cross-linguistic similarity, the manner in which these dynamics may change over the course of L2 learning, and the influence of a changing construct of perceived (dis)similarity in shaping L2 outcomes (cf. Aoyama et al., 2004) remain open questions.

Linking L2 perception and L2 production

Although theories of L2 speech such as the SLM are often concerned with fundamental similarities between child and adult learners, there are some important differences between L1 learning and L2 learning of a target language, and one area in which such differences are observed is the link between perception and production. According to NLM-e, perception and production are closely linked in L1 acquisition, and these links are understood to form during L1 acquisition, due in part to the articulatory-auditory loop associated with an infant hearing the consequences of her own vocalizations. Further, the timing of perception and production milestones in L1 development, which typically shows children reliably perceiving speech sounds well before they can produce them, suggests that perception generally precedes production when it comes to the L1 (although, given the aforementioned role of production in perceptual development, this should be regarded as a gross view of perception-production ordering in L1 development).

In L2 speech learning, the relationship between perception and production is less clear than in L1 development. One body of findings that bears on this issue comes from research on phonetic training – in particular, transfer of training gains from the trained modality to the untrained modality. In brief, although some studies report perception-production correlations and/or carryover of training gains across modalities (e.g., Catford and Pisoni, 1970; Leather, 1996; Wang, Jongman, and Sereno, 2003; Kartushina, Hervais-Adelman, Frauenfelder, and Golestani, 2015), much of this literature evinces little to no relationship between perception and production for L2 speech. For example, Kartushina and Frauenfelder (2014) tested a sample of L1 Spanish learners of French on their perception and production of French front mid vowel contrasts and found no correlation between learners' performance in perception and production (see also Peperkamp and Bouchon, 2011). Of course, failure to find a statistically significant correlation between the two modalities does not constitute evidence that there is no relationship between them; however, when correlations have been found across modalities, the effect size has often been small (Bradlow et al., 1997; Akahane-Yamada, McDermott, Adachi, Kawahara, and Pruitt, 1998; Flege, MacKay, and Meador, 1999). Furthermore, L2 production accuracy does not seem to depend on high L2 perceptual ability (Sheldon and Strange, 1982). These findings suggest that transfer of perceptual learning to production, and vice versa, may be complicated by a variety of intervening factors (for recent reviews, see Kartushina, 2015; Kartushina et al., 2015).

The complex relationship between L2 developments in perception and production invites the question of whether L2 outcomes might generally benefit from training involving more than one modality (see Chapter 19, this volume, discussing multichannel training of the L2). If L2 training in one modality tends to improve mainly the trained modality and not necessarily the untrained modality, could L2 speech development be enhanced or accelerated with multimodal training involving more than one stimulus channel (e.g., auditory and articulatory, auditory and visual)? The logic of "more is better" would predict yes, and some findings show a significant benefit of visual feedback on learners' production combined with auditory exposure to a native model (e.g., Kartushina, 2015). However, two recent studies suggest that multimodal L2 engagement does not necessarily improve L2 outcomes, and in certain cases can actually be detrimental to L2 speech development. In one study of L1 English speakers being trained on Mandarin tones (Godfroid, Lin, and Ryu, 2017), several types of multimodal perceptual training were systematically compared to each other, including three "single-cue" types of training (involving exposure to only one visual cue alongside auditory stimuli: number, pitch contour, or color) and two "dual-cue" types of training (involving exposure to two visual cues: color and number, or color and pitch contour). Although test results showed perceptual gains with all training types, single-cue exposure to numbers or pitch contours was more beneficial than single-cue exposure to colors, while neither dual-cue exposure was more beneficial than single-cue exposure. In another study of L1 Spanish learners of Basque (Baese-Berk and Samuel, 2016), both inexperienced and experienced learners were trained on L2 Basque sounds in two conditions: (a) perception only and (b) perception and production (where an oral repetition task was interleaved with auditory exposure and perceptual judgments). Results showed a detrimental effect of producing speech during the perceptual training, which was mitigated (but not eliminated) by previous experience with the L2 (see also Leach and Samuel, 2007; Baese-Berk, 2010, for similar findings).

Why would multimodal L2 exposure not necessarily facilitate L2 learning outcomes? There are at least three possible, and not mutually exclusive, explanations. Baese-Berk and Samuel (2016) allude to the role of cognitive (over)load in this type of multimodal setting (see, e.g., van Merriënboer and Sweller, 2005), while Godfroid et al. (2017) additionally consider the potentially detrimental effect of extraneous processing of redundant or irrelevant information. Either or both of these factors may be responsible for the observed interference associated with varying task-irrelevant phonetic features in L2 speech training (Antoniou and Wong, 2016). Godfroid et al. (2017), however, are careful to point out that performance in their dual-cue condition was never *worse* than in the single-cue condition, which is not entirely consistent with an account of their results in terms of cognitive load or extraneous processing. Consequently, they explain the lack of benefit associated with adding a second cue in terms of an implementation issue: given the way in which color was incorporated into the stimuli, participants may have perceptually backgrounded the color cue, such that “color might have played a more peripheral role than intended” (Godfroid et al., 2017, p. 846).

Thus, the findings of L2 speech research present a mixed picture regarding the relationship of perception and production in L2 development. On the one hand, studies such as Bradlow et al. (1997) and Baese-Berk and Samuel (2016), which show facilitation or interference across modalities, provide evidence that perception and production processes must draw on mental representations that are at least partly shared between the two modalities. On the other hand, studies such as Kartushina and Frauenfelder (2014), which fail to find a close correlation between perception and production performance in an L2, suggest some degree of dissociation between perception and production representations as well. The degree to which L2 perception and L2 production processes overlap, the nature of this overlap, and the manner in which the perception-production relationship differs between L1 and L2 learning remain some of the basic questions in research on L2 speech learning and phonetic development more generally.

Concluding remarks

This chapter has attempted to synthesize core insights and claims of influential theories in the field of L2 speech learning. The four selected frameworks discussed here (PAM-L2, NLM, ASP, SLM) are some of the most detailed and widely tested theories in the field. However, it should be noted that other frameworks, such as Eckman’s Markedness Differential Hypothesis and Structural Conformity Hypothesis (Eckman, 1977, 1991, 2008), Brown’s featural model of L2 perception (Brown, 2000), the Ontogeny Phylogeny Model (Major, 2001), the Second Language Linguistic Perception Model (Escudero, 2009; van Leussen and Escudero, 2015), and additional Optimality Theoretic approaches (cf. Hancin-Bhatt, 2008), also address aspects of L2 speech learning with different emphases, such as the role of language-universal factors (Markedness Differential Hypothesis) and developmental changes over the course of learning (Ontogeny Phylogeny Model). Furthermore, the burgeoning field of L2 prosody has led to new developments and theories focusing on suprasegmental features of the L2, such as rhythm and intonation (e.g., Li and Post, 2014; Mennen and de Leeuw, 2014; Mennen, 2015).

In closing, although the focus of this chapter has been phonetic development in typical late-onset L2 learners, it is worth drawing the reader's attention to some related areas of research activity that, for reasons of space, have not been given extensive discussion here. First, CLI is being understood in new ways, not just in terms of "negative" or "positive" transfer from the L1 (Lado, 1957; Goto, 1971; Odlin, 1989; Cutler, 2001) but also in terms of native-language transfer benefits for L2 learning (Bohn and Best, 2012; Chang and Mishler, 2012; Chang, 2016). Moreover, the variables of age and profile of acquisition have long spurred, and continue to spur, research on differences between early and late L2 learners (Yeni-Komshian et al., 2000; Guion, 2003; Kang and Guion, 2006; Oh et al., 2011) and between typical L2 learners and heritage speakers or L1 (re)learners (Knightly, Jun, Oh, and Au, 2003; Chang, 2011; Chang and Yao, 2016). Additional work has been examining the effect of other properties of the individual learner, such as language aptitude and basic perceptual ability, in order to better understand the wide range in L2 outcomes observed across learners of the same L1 background (e.g., Chang, 2010b; Perrachione, Lee, Ha, and Wong, 2011; Bowles, Chang, and Karuzis, 2016). Input and linguistic factors such as talker variability and phonological context are also being investigated as contributors to variation in L2 outcomes (e.g., Bradlow et al., 1997; Kingston, 2003; Perrachione et al., 2011; Chang and Bowles, 2015).

In light of transnational migration and multilingualism across the world, these and other lines of inquiry are poised to shape practices and policy affecting the linguistic lives of many people. For example, research examining the linguistic knowledge of heritage speakers in relation to late L2 learners is helping to inform language course design so as to better serve the unique needs of heritage language learners. Studies of third language (L3) learners are investigating what factors influence speech learning in multilingual situations (Gallardo del Puerto, 2007), how L3 learning resembles and differs from L2 learning (Onishi, 2013; Wrembel, 2014), and how L3 learning may influence the phonological representations and processes associated with previously learned languages (Cabrelli Amaro, 2017), in line with the multicompetence view of language development over the lifespan (Cook, 1991, 1992, 2003). Finally, it would be remiss not to mention the work of scholars who are connecting L2 speech research to L2 instructional practices and strategies (see, e.g., Mora and Levkina, 2017). Far from declining, research in L2 speech learning is thriving. We can look forward to many new discoveries in the years to come, with practically relevant implications we cannot yet imagine.

Notes

1 This contrasts with the direct realist view of PAM(-L2), which assumes that perception-production links do not have to be learned. Under this view, both perception and production are based in articulatory gestures, so "no translation is needed between perception and production because they are informationally compatible" (Best, 1994, p. 180).

2 Thanks to Geoffrey Schwartz for pointing out that different outcomes for acoustic proximity could result from examining alternative metrics of vowel quality (e.g., distances between adjacent formants such as F_2 and F_3; cf. Syrdal and Gopal, 1986, on the role of formant convergences in vowel perception). In the case of measuring vowel frontness/backness in terms of F_3-F_2 (rather than F_2), this metric, when applied to the production data from the speakers in Chang et al. (2011), is consistent with the view that native English /u/ (M_{F3-F2} = 3.617 Bark) is phonetically closer to native Mandarin /y/ (M_{F3-F2} = 1.367 Bark) than to native Mandarin /u/ (M_{F3-F2} = 8.588 Bark).

References

Ahn, S., Chang, C. B., DeKeyser, R., and Lee-Ellis, S. (2017). Age effects in first language attrition: Speech perception by Korean-English bilinguals. *Language Learning*, *67*(3), 694–733.

Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., and Pruitt, J. S. (1998). Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP-1998)* (paper 0429). Sydney, Australia: International Speech Communication Association.

Altenberg, E. P. (2005a). The judgment, perception, and production of consonant clusters in a second language. *International Review of Applied Linguistics in Language Teaching*, *43*(1), 53–80.

Altenberg, E. P. (2005b). The perception of word boundaries in a second language. *Second Language Research*, *21*(4), 325–358.

Antoniou, M., Best, C. T., Tyler, M. D., and Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*(4), 640–653.

Antoniou, M., and Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America*, *139*(1), 271–278.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250.

Baese-Berk, M. M. (2010). *The relationship between perception and production when learning novel phonological categories*. Ph.D. thesis, Northwestern University.

Baese-Berk, M. M., and Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman and 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. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore, MD: York Press.

Best, C. T., McRoberts, G. W., and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794.

Best, C. T., McRoberts, G. W., and 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*(3), 345–360.

Best, C. T., Traill, A., Carter, A., Harrison, K. D., and Faber, A. (2003). !Xóõ click perception by English, Isizulu, and Sesotho listeners. In M. J. Solé, D. Recasens, and J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 853–856). Barcelona, Spain: Causal Productions.

Best, C. T., and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn and M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins Publishing.

Bohn, O. S. (2017). Cross-language and second language speech perception. In E. M. Fernández and H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 213–239). Hoboken, NJ: John Wiley & Sons.

Bohn, O. S., and Best, C. T. (2012). Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics*, *40*(1), 109–128.

Boomershine, A., Hall, K. C., Hume, E., and Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, E. Dresher, and K. Rice (Eds.), *Contrast in phonology* (pp. 143–172). Berlin, Germany: Mouton de Gruyter.

Bowles, A. R., Chang, C. B., and Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*(4), 774–808.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/, IV: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310.

Broselow, E., and Kang, Y. (2013). Phonology and speech. In J. Herschensohn and M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 529–554). Cambridge: Cambridge University Press.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Malden, MA: Blackwell Publishers.

Cabrelli Amaro, J. (2017). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism*, *21*(6), 698–717.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., and Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421–428.

Catford, J. C., and Pisoni, D. B. (1970). Auditory vs. Articulatory training in exotic sounds. *The Modern Language Journal*, *54*(7), 477–481.

Celata, C. (in press). Phonological L1 attrition. In M. S. Schmid and B. Köpke (Eds.), *The Oxford handbook of language attrition*. Oxford: Oxford University Press.

Chang, C. B. (2009). English loanword adaptation in Burmese. *Journal of the Southeast Asian Linguistics Society*, *1*, 77–94.

Chang, C. B. (2010a). *First language phonetic drift during second language acquisition*. Ph.D. thesis, University of California, Berkeley, CA.

Chang, C. B. (2010b). The implementation of laryngeal contrast in Korean as a second language. *Harvard Studies in Korean Linguistics*, *13*, 91–104.

Chang, C. B. (2011). Systemic drift of L1 vowels in novice L2 learners. In W. S. Lee and E. Zee (Eds.), *Proceedings of the 17th international congress of phonetic sciences* (pp. 428–431). Hong Kong, China: City University of Hong Kong.

Chang, C. B. (2012a). Phonetics vs. Phonology in loanword adaptation: Revisiting the role of the bilingual. In S. Berson, A. Bratkievich, D. Bruhn, A. Campbell, R. Escamilla, A. Giovine, L. Newbold, M. Perez, M. Piqueras-Brunet, and R. Rhomieux (Eds.), *Proceedings of the 34th annual meeting of the Berkeley Linguistics Society: General session and parasession on information structure* (pp. 61–72). Berkeley, CA: Berkeley Linguistics Society.

Chang, C. B. (2012b). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249–268.

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 491–504.

Chang, C. B. (2015). Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In E. Raimy and C. E. Cairns (Eds.), *The segment in phonetics and phonology* (pp. 199–217). Chichester: John Wiley & Sons.

Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, *19*(4), 791–809.

Chang, C. B. (2018). Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics*, *68*, 85–102.

Chang, C. B. (in press). Phonetic drift. In M. S. Schmid and B. Köpke (Eds.), *The Oxford handbook of language attrition*. Oxford: Oxford University Press.

Chang, C. B., and Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *The Journal of the Acoustical Society of America*, *138*(6), 3703–3716.

Chang, C. B., and Mishler, A. (2012). Evidence for language transfer leading to a perceptual advantage for non-native listeners. *The Journal of the Acoustical Society of America*, *132*(4), 2700–2710.
Chang, C. B., and Yao, Y. (2016). Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal*, *13*(2), 134–160.
Chang, C. B., Yao, Y., Haynes, E. F., and Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America*, *129*(6), 3964–3980.
Colantoni, L., Steele, J., and Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge: Cambridge University Press.
Cook, V. J. (1991). The poverty-of-the-stimulus argument and multicompetence. *Second Language Research*, *7*(2), 103–117.
Cook, V. J. (1992). Evidence for multicompetence. *Language Learning*, *42*(4), 557–591.
Cook, V. J. (2003). The changing L1 in the L2 user's mind. In V. Cook (Ed.), *Effects of the second language on the first* (pp. 1–18). Clevedon: Multilingual Matters.
Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, *5*(1), 1–23.
Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*(1), 104–137.
Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 270–282.
Davidson, L. (2017). Cross-language speech perception and production. In M. Aronoff (Ed.), *Oxford bibliographies in linguistics*. Oxford: Oxford University Press.
de Bot, K., Lowie, W., and Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7–21.
de Leeuw, E. (in press). Phonetic L1 attrition. In M. S. Schmid and B. Köpke (Eds.), *The Oxford handbook of language attrition*. Oxford: Oxford University Press.
de Leeuw, E., Mennen, I., and Scobbie, J. M. (2013). Dynamic systems, maturational constraints and L1 phonetic attrition. *International Journal of Bilingualism*, *17*(6), 683–700.
de Leeuw, E., Schmid, M. S., and Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(1), 33–40.
Dmitrieva, O., Jongman, A., and Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics*, *38*(3), 483–492.
Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578.
Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315–330.
Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition*, *13*(1), 23–41.
Eckman, F. R. (2008). Typological markedness and second language phonology. In J. G. Hansen Edwards and M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 95–115). Amsterdam, The Netherlands: John Benjamins Publishing.
Eckman, F. R. (2012). Second language phonology. In S. M. Gass and A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 91–105). New York, NY: Routledge.
Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194.
Escudero, P. (2009). Linguistic perception of "similar" L2 sounds. In P. Boersma and S. Hamann (Eds.), *Phonology in perception* (pp. 151–190). Berlin, Germany: Mouton de Gruyter.
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*(1), 47–65.

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. 233–272). Baltimore, MD: York Press.

Flege, J. E. (1996). English vowel productions by Dutch talkers: More evidence for the “similar” vs “new” distinction. In A. James and J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 11–52). Berlin, Germany: Mouton de Gruyter.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, and A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217–244). Trier, Germany: Wissenschaftlicher Verlag.

Flege, J. E., and Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185–202.

Flege, J. E., and Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*(1), 67–83.

Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). Native Italian speakers’ perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987.

Fokes, J., Bond, Z. S., and Steinberg, M. (1985). Acquisition of the English voicing contrast by Arab children. *Language and Speech*, *28*(1), 81–92.

Gallardo del Puerto, F. (2007). Is L3 phonological competence affected by the learner’s level of bilingualism? *International Journal of Multilingualism*, *4*(1), 1–16.

Gass, S. (1984). Development of speech perception and speech production in adult second language learners. *Applied Psycholinguistics*, *5*(1), 51–74.

Godfroid, A., Lin, C. H., and Ryu, C. (2017). Hearing and seeing tone through color: An efficacy study of web-based, multimodal Chinese tone perception training. *Language Learning*, *67*(4), 819–857.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, *9*(3), 317–323.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, *60*(2), 98–128.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America*, *102*(1), 655–658.

Hallé, P. A., and Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *The Journal of the Acoustical Society of America*, *121*(5), 2899–2914.

Hancin-Bhatt, B. (2008). Second language phonology in optimality theory. In J. G. Hansen Edwards and M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 117–146). Amsterdam, The Netherlands: John Benjamins Publishing.

Hazan, V. (2007). Second language acquisition and exemplar theory. In J. Trouvain and W. J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences* (pp. 39–42). Dudweiler, Germany: Pirrot.

Holliday, J. J. (2016). Second language experience can hinder the discrimination of nonnative phonological contrasts. *Phonetica*, *73*(1), 33–51.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson and J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Kang, K. H., and Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *The Journal of the Acoustical Society of America*, *119*(3), 1672–1683.

Kang, Y. (2008). Interlanguage segmental mapping as evidence for the nature of lexical representation. *Language and Linguistics Compass*, *2*(1), 103–118.

Kartushina, N. (2015). *Second language phonological acquisition in adults*. Ph.D. thesis, University of Geneva.

Kartushina, N., and Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, 1246.

Kartushina, N., Frauenfelder, U. H., and Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning*, *66*(S2), 155–186.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319.

Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, *46*(2–3), 295–349.

Knightly, L. M., Jun, S. A., Oh, J. S., and Au, T. K. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America*, *114*(1), 465–474.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*(2), 93–107.

Kuhl, P. K, Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, *363*, 979–1000.

Kuhl, P. K., and Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore, MD: York Press.

LaCharité, D., and Paradis, C. (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry*, *36*(2), 223–258.

Ladefoged, P. (1999). American English. In The International Phonetic Association (Ed.), *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet* (pp. 41–44). Cambridge: Cambridge University Press.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.

Laeufer, C. (1996). Towards a typology of bilingual phonological systems. In A. James and J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 325–342). Berlin, Germany: Mouton de Gruyter.

Leach, L., and Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353.

Leather, J. (1996). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In A. James and J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 75–101). Berlin, Germany: Mouton de Gruyter.

Levy, E. S., and Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36*(1), 141–157.

Li, A., and Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition*, *36*(2), 223–255.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception and Psychophysics*, *46*(2), 187–200.

Mack, M. (1990). Phonetic transfer in a French-English bilingual child. In P. H. Nelde (Ed.), *Language attitudes and language conflict* (pp. 107–124). Bonn, Germany: Dümmler.

Mack, M. (2003). The phonetic systems of bilinguals. In M. T. Banich and M. Mack (Eds.), *Mind, brain, and language: Multidisciplinary perspectives* (pp. 309–349). Mahwah, NJ: Lawrence Erlbaum Press.

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, *76*(2), 190–208.

Major, R. C. (1996). L2 acquisition, L1 loss, and the critical period hypothesis. In A. James and J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 147–159). Berlin, Germany: Mouton de Gruyter.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.
Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussarie, M. Avanzi, and S. Herment (Eds.), *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations* (pp. 171–188). Berlin, Germany: Springer Verlag.
Mennen, I., and de Leeuw, E. (2014). Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition*, *36*(2), 183–194.
Mora, J. C., and Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, *39*(2), 381–399.
Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, *39*(2), 156–167.
Onishi, H. (2013). *Cross-linguistic influence in third language perception: L2 and L3 perception of Japanese contrasts*. Ph.D. thesis, University of Arizona.
Peperkamp, S., and Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of the 12th international conference on spoken language processing (Interspeech-2011)* (pp. 168–171). Florence, Italy: International Speech Communication Association.
Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472.
Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam, The Netherlands: John Benjamins Publishing.
Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic and acoustic contributions. *The Journal of the Acoustical Society of America*, *89*(6), 2961–2977.
Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics*, *52*(1), 37–52.
Polka, L., and Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, *100*(1), 577–592.
Port, R. F., and Mitleb, F. M. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics*, *11*(3), 219–229.
Schwartz, G. (2016). Word boundaries in L2 speech: Evidence from Polish learners of English. *Second Language Research*, *32*(3), 397–426.
Sheldon, A., and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261.
Simonet, M. (2016). The phonetics and phonology of bilingualism. In *Oxford handbooks online* (pp. 1–25). Oxford: Oxford University Press.
Strange, W. (2011). Automatic Selective Perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466.
Strange, W., Levy, E., and Lehnholf, R., Jr. (2004). Perceptual assimilation of French and German vowels by American English monolinguals: Acoustic similarity does not predict perceptual similarity. *The Journal of the Acoustical Society of America*, *115*(5), 2606.
Syrdal, A. K., and Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, *79*(4), 1086–1100.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., and Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, *123*(1), 397–413.

Tice, M., and Woodley, M. (2012). Paguettes and bastries: Novice French learners show shifts in native phoneme boundaries. *UC Berkeley Phonology Lab Annual Report*, 72–75.

Tobin, S. J., Nam, H., and Fowler, C. A. (2017). Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics*, *65*, 45–59.

Tucker, G. R. (2001). A global perspective on bilingualism and bilingual education. In J. E. Alatis and A. H. Tan (Eds.), *Language in our time: Bilingual education and official English, Ebonics and standard English, immigration and the Unz initiative* (pp. 332–340). Washington, DC: Georgetown University Press.

Tyler, M. D., Best, C. T., Faber, A., and Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4–21.

van Leussen, J. W., and Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000.

van Merriënboer, J. J. G., and Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177.

Wang, Y., Jongman, A., and Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658.

Weinreich, U. (1953). *Languages in contact: Findings and problems*. The Hague, The Netherlands: Mouton de Gruyter.

Werker, J. F., and Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics*, *21*(4), 289–297.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics*, *26*(2), 95–104.

Wrembel, M. (2014). VOT patterns in the acquisition of third language phonology. *Concordia Working Papers in Applied Linguistics*, *5*, 750–770.

Yeni-Komshian, G. H., Flege, J. E., and Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, *3*(2), 131–149.

Zsiga, E. (2011). External sandhi in a second language: The phonetics and phonology of obstruent nasalization in Korean and Korean-accented English. *Language*, *87*(2), 289–345.