The Cambridge Handbook of Second Language Acquisition

Phonology and phonetics

Ellen Broselow and Yoonjung Kang

Abstract

Learners of a new language must acquire a new inventory of sound contrasts, new restrictions on where sounds may occur, and new prosodic structures. The challenges facing researchers are to understand the characteristics of L2 speech and to explain how and why those characteristics arise. This chapter provides an overview of research on how speakers acquire the sound system of a new language, focusing on three major issues: (i) the extent to which L2 sound patterns are influenced by transfer from the first language and/or by linguistic universals; (ii) the level of representation (lexical, phonemic, phonetic, featural) at which L2 acquisition occurs; and (iii) the relationship between the perception and production of the sound system of a new language. These questions are explored with reference to the acquisition of segmental contrasts, restrictions on the occurrence of consonants in different syllabic positions, and prosodic structures (tone, pitch accent, stress, and intonation). The chapter concludes with a discussion of recent advances in methodology and in grammar modeling, and outlines directions for future study.

Keywords: second language phonology, second language phonetics, phonemes, phonotactics, syllable structure, prosody, perception, contrast, markedness, universals

21.1 Introduction

The acquisition of the sound system of a new language poses significant challenges for learners, who must acquire a new inventory of sound contrasts, new restrictions on where sounds may occur, and new prosodic structures. Researchers also face challenges: to understand the characteristics of L2 speech and to explain how and why those characteristics arise. This chapter provides an overview of research on how speakers acquire the sound system of a new language, focusing on three major issues: the extent to which L2 sound patterns are influenced by transfer from the first language and/or by linguistic universals; the level of representation (lexical, phonemic, phonetic, featural) at which L2 acquisition occurs; and the relationship between the perception and production of the sound system of a new language. These questions are explored with reference to the acquisition of contrasts in stops and vowels, restrictions on the occurrence of consonants in different syllabic positions, and prosodic structures of tone, pitch accent, stress, and intonation. The chapter concludes with a discussion of recent advances in methodology and in grammar modeling, and outlines directions for future study.

21.1.1 L1 Transfer and the Influence of Universal Markedness

Almost all studies on the acquisition of L2 sound systems address the question of how L1 influence contributes to shaping L2 learners' sound patterns. Lado (1957) predicted that those aspects of L2 that are similar to the L1 will be easily acquired, while those aspects that are different from the L1 will pose greater difficulty. Aside from the challenge of defining what counts as "similar", two recurring phenomena challenge the claim that L2 errors are based solely on the difference between the L1 and the L2. First, certain L2 structures not present in L1 appear to be more easily acquired than others. Second, the speech of L2 learners in many cases exhibits patterns that are coherent and systematic, but are nonetheless distinct from those of both the target L2 and the native language. For example, learners whose L1 lacks any final obstruents may exhibit an interlanguage that contains voiceless but not voiced final obstruents, though both structures are equally novel for the learner and are equally present in the L2 input (Section 21.3.2). Phenomena of both types have frequently been attributed to universal principles such as markedness, where unmarked structures are generally considered to be those that are more basic, typologically common, and phonetically easier to perceive and produce than more marked structures (see Rice, 2007); such facts led Eckman (1977) to propose that marked L2 structures are more difficult to acquire than equally novel but less marked structures (see Archibald, 2021 for a review of approaches to this problem).

Markedness has been encoded in theories of grammar in various ways. In a framework that views a grammar as a set of parameters with different possible settings (e.g., a parameter allowing consonant clusters in syllable onsets, which may be set either on or off), the preference for less marked structures can be viewed as encoded in default parameter settings (e.g., the default setting of "no" for onset clusters: see Blevins, 1995). In a framework that views the grammar as a set of ranked, universal constraints. e.g., Optimality Theory (OT) as defined by Prince and Smolensky (2004), markedness is encoded as a set of structural constraints which are assumed to be present in the grammars of all languages, but may be rendered inactive by more highly ranked constraints; for example, a constraint banning consonant clusters in onsets may be outranked by a constraint demanding that lexical forms be realized "faithfully." Exposure to L2 data may trigger resetting of parameters (e.g., Archibald, 1993, 1994) or rerankings of constraints (e.g., Eckman, 2004; Hancin-Bhatt, 2000).

That a learner's previous language experience affects their learning of new languages is uncontroversial; such effects have been found even across language modalities, in the learning of signed languages by hearing speakers (e.g., Ortega & Morgan, 2015). Yet as the cases discussed in this chapter illustrate, teasing apart L1 effects from the effects of universal tendencies is not trivial. Complicating this issue is the tendency to map L2 sounds to similar but not necessarily identical L1 phoneme categories, recognized in a number of models: the Speech Learning Model (Flege, 1995) and its revision (Flege & Bohn, 2021), where it is termed "equivalence classification"; in the Perceptual Assimilation Model (Best, 1995) and its extension to L2 acquisition (Best & Tyler, 2007); and in the Native Language Magnet theory (Kuhl & Iverson, 1995). Each of these models outlines the ways in which this process may interfere with the accurate acquisition of L2 sounds– for example, by mapping contrasting sounds in the L2 to a single L1 category– though the degree of L1 interference may be modulated by factors such as differences in experience with the L2, age of L2 exposure, speech rate, attention and language mode in experimental conditions, and cognitive abilities (Antoniou et al., 2010; Darcy, Mora, & Diadone, 2016; Piske, MacKay, & Flege, 2001).

21.1.2 Phonetics, Phonology, and the Lexicon

A second major issue in the acquisition of L2 sound systems concerns whether specific aspects of L2 sound patterns are best explained from the perspective of phonology (an internalized system of abstract rules or constraints defining the possible sound patterns of a language) or of phonetics (the physical implementation of speech sounds in production and the interpretation of fine-grained acoustic cues in perception); see Davidson (2011) for a review. The boundary between phonetics and phonology, however, is increasingly blurred in recent literature, with no clear consensus on where the boundary should be drawn if at all (Browman & Goldstein, 1989; Davidson, 2006; Escudero & Boersma, 2004; Zsiga, 2003). Various accounts have assumed that the reason learners fail to produce L2 structures correctly is because their phonological grammar does not sanction them, causing L2 forms to be modified by the L1 grammar. In this approach, L2 acquisition involves moving from the the L1 grammar toward an L2-like grammar in response to L2 input, which may trigger (depending on one's model of grammar) resetting of parameters, changes in phonological rules (Eckman et al., 2003; Young-Scholten 2004), or changes in the ranking of constraints (e.g., Escudero & Boersma, 2004; Hancin-Bhatt, 2000). On the other hand, certain L2 modifications are argued to be better explained at the level of phonetics, where speakers may simply fall short in mastering the correct articulation (Colantoni & Steele, 2007; Davidson, 2010) or fail to correctly perceive the L2 target forms, which, in turn, can lead to production modification (e.g., Flege, 1995).

A related issue concerns the units of representation involved in L2 acquisition (Kang, 2008) – that is, whether the L2 is perceived and produced in terms of abstract phonological units such as phonemes or phonological distinctive features (e.g., Brown, 2000; Hancin-Bhatt, 1994; Larson-Hall, 2004; Yazawa et al., 2023); in terms of surface phonetic features (Brannen 2002; de Jong, Hao & Park, 2009; de Jong, Silbert & Park, 2009); or in terms of more phonetically based notions such as articulatory gestures (Best & Halle, 2010). Researchers beginning from the perspective of phonology have proposed that the attainability of L2 structures is a function of the availability of corresponding phonological structures in the L1 (Archibald, 2005; Brown, 2000; Goad & White, 2006). Distinctive features (which serve both to distinguish contrasting phonemes of the language and to organize sounds into natural classes sharing a particular feature) have been appealed to for explanations of crosslinguistic differences in L2 perception as well as in phoneme substitution. For example, Hancin-Bhatt (1994) argued that contrastive phonological features of L1 mediate the choice of L1 phoneme to replace an illegal L2 sound (e.g., why English /θ/ is typically replaced with /t/ by speakers of Turkish but with /s/ by speakers of Japanese, even though /t/ and /s/ are present in both L1 phoneme inventories). Investigating the differential substitution patterns for English interdental fricatives by speakers of German, Hindi, Japanese and Turkish, Hancin-Bhatt proposed that the likeliest L1 substitute is one that preserves those features of the L2 sound which carry the highest prominence in the L1, with feature prominence being defined quantitatively by the number of contrasts inventory the feature encode in the L1 phoneme.

However, other studies have suggested that at least some L2 patterns must be described in terms of surface phonetic features rather than contrastive phonological features. According to Brannen (2002), the English interdental fricative /θ/ is replaced by /s/ in European French and by /t/ in Quebec French, although the two dialects have identical phoneme systems because the coronal fricative /s/ is dental in European French, but alveolar in Quebec French, a phonetic detail that is not considered contrastive in either dialect. Similarly, Chang (2018) proposes that L2 perception responds to the informativeness of an acoustic cue in the L1 rather than the L1 phonological status of the feature signaled by the cue, based on the finding that L1 Korean speakers outperformed native English speakers in the identification and detection of English final stops without release burst (e.g., *whea[**t̚**] vs. wee[**p̚**]* or *bee[**t̚**]* vs. *bee*). Chang attributes this to the fact that in Korean, word-final stops (unlike English stops in this position) are obligatorily realized without a release burst, making the vocalic transitions the most informative cue to consonant place for Korean (but not English) speakers.

Evidence suggesting that both phonological features and the mastery of articulatory routines play a role in L2 acquisition comes from comparing the rate at which contrasts defined by the same phonological feature are acquired. De Jong, Silbert and Park (2009) and de Jong, Hao and Park (2009) found that Korean learners' ability to perceive the English stop–fricative contrast for both labial /b–v/ and coronal /d–ð/ proceeded similarly, suggesting that the learning of one contrast defined by the feature [continuant] generalized to continuant contrasts across all places of articulation; in production, however, the accuracy of the continuancy contrasts in labial and coronal place did not correlate, presumably because implementation of the two contrasts involves different articulatory routines. There was, however, a correlation between the production accuracy of learners' stop–fricative contrasts in voiced labials /b–v/ and voiceless labials /p–f/, since in this case the gestures for producing the manner contrast are similar across the voiced and voiceless pairs. Yazawa et al. (2023) found support for feature-mediated perception of novel English vowels by Japanese speakers. Japanese listeners were able to form a new vowel category (/æ/) when they could combine existing L1 features in a new way (*/low, front/), but were not able to form new categories for English vowels (e.g., /ɛ/) that map to existing feature combinations of Japanese vowels (e.g., /mid, front/ = Japanese /e/). They model these results in the Second Language Linguistic Perception framework (Escudero 2009), in which a series of cue constraints guide the mapping from acoustic cues to sound categories, and L2 learning is error-driven adjustment of constraint ranking A feature-based model successfully reflected the experimental results, while a segment-based model was not able to account for the asymmetry in acquisition of equally novel vowel categories.

Flege (1995) proposed that segmental acquisition takes place on the basis of allophones (context-specific realizations of phonemes) at a level more abstract than surface phones but less abstract than phonemes. This view is supported by numerous studies showing that a given L2 segment is neither perceived nor produced with uniform accuracy across different contexts (e.g., onset vs. coda). Some attribute such positional variation to phonological markedness, while others seek explanations based on perceptual similarity, articulatory ease, and L2 input frequency. For example, Trofimovich, Gatbonton and Segalowitz (2007) showed that the production of English /ð/ by French speakers was more accurate in some positions (e.g., sentence-initial) than in others, concluding that the target sound is more easily learned in positions where it is more frequent and where it is perceptually more distinct from L1 sounds.

Evidence for position-specific acquisition is also found in perception. For example, Ingram and Park (1998) found that Korean speakers were better at perceiving the English /l/–/ɹ/ contrast in intervocalic or cluster position, where the English contrast can be equated with the native Korean contrast between singleton [ɾ] vs. geminate liquid [ll], than in initial position, where no singleton vs. geminate contrast is available in Korean. Japanese speakers, on the other hand, who have no comparable singleton vs. geminate liquid contrast in L1, showed generally poorer perception of the English /l/–/ɹ/ contrast than Korean speakers, but the Japanese speakers' perception was better in intervocalic and initial positions, where acoustic cues are generally more perceptible, than in clusters. They conclude that L1 background and general acoustic discriminability, but not universal markedness, affected the perception pattern.

Others present a dynamic view of the level of L2 sound processing. Best and Tyler (2007) propose that unlike naïve listeners, who perceive unfamiliar language sounds at the phonetic level, L2 listeners may access both the phonological and the phonetic levels of L2 sounds. In Strange's (2011) Automatic Selective Perception model, beginning learners of L2 employ a "phonetic mode of perception," while advanced L2 learners develop "selective perception routines" that identify

L2 phonological categories. Similarly, Chang (2015, p. 204) proposes that "relatively experienced L2 users are expected to show L1-L2 mappings that follow phonemic similarity over acoustic and allophonic similarity because of a tendency for high-level information to override low-level information."

A debate also exists regarding the role of top-down lexical knowledge in acquiring L2 contrasts. Van Leussen and Escudero (2015) present a simulation of L2 category acquisition in which L2 lexical contrasts guide the establishment and fine-tuning of L2 categories from the initial L1-like state. Hayes-Harb (2007) found that lexical contrasts facilitate acquisition of a new contrast in an artificial language learning experiment, where English speakers performed better in distinguishing pseudo-words that contained voiced [g] and voiceless unaspirated [k] (which do not normally serve a contrastive function in English) when stimuli were presented with contrastive meanings ([ga] "pot" vs. [ka] "mouse"). Relatedly, Bundgaard-Nielsen, Best, and Tyler (2011) found that L1 Japanese/L2 Australian English participants' ability to consistently map L2 vowels to matching L1 categories increased with vocabulary size, suggesting that L2 learners attune their L2 perception through lexical acquisition.

However, perceptual discrimination of L2 contrasts cannot always always be equated with lexical encoding. Hayes-Harb and Masuda (2008) found that some English speakers were successful at discriminating the singleton–geminate consonant contrast in made-up Japanese words (e.g., *meso* "piano" vs. *messo* "boot"), but were unable to implement this contrast in production, presumably due to the failure to establish an accurate lexical representation of the (correctly perceived) short–long contrast. Similarly, in a priming task, Broersma (2012) found differences in lexical activation and inhibition of particular L2-specific contrasts between L1 and less advanced L2 speakers, suggesting that even when listeners are able to perceive L2 contrasts, their lexical encoding and access may not be nativelike.

Other studies found evidence of lexical encoding in the absence of reliable perception. Using eye-tracking technology, Cutler, Weber, and Otake (2006) found that Japanese learners of English hearing English *rock* looked both at pictures of a rock and its minimal pair competitor, a lock, although when the participants heard *lock*, they converged on the lock picture. Cutler et al. interpreted this asymmetrical pattern as evidence that the learners had encoded the /l/ vs. /ɹ/ contrast lexically but could not perceive the contrast reliably though an eye tracking study of L1 Dutch speakers' perception of English vowel contrasts by Escudero, Hayes-Harb, and Mitterer (2008) suggested that such asymmetries may be driven by orthographic information. Darcy et al. (2012) also report a case of lexical contrast without perceptual contrast: in a priming task, their advanced English-speaking learners of French exhibited sensitivity to lexical distinctions for French front vs. back rounded vowels, although their perceptual discrimination of the same contrast was less accurate than native French speakers. Relatedly, a lexical decision task probing lexical encoding of L2 Spanish contrasts (e.g., /ɾ/ vs. /d/) by English speaking learners revealed vocabulary size to be a more effective predictor of accurate lexical encoding than perceptual ability (Daidone & Darcy, 2021).

Nonetheless, sensitivity to novel contrasts may develop in the absence of lexical knowledge. Distributional learning, whereby infants learn to discriminate speech sounds based on passive exposure to relative frequency distributions in the continuous auditory signal in a laboratory setting (Maye, Werker, & Gerken, 2022), was found effective in adult Spanish learners' acquisition of the Dutch /a:/ vs. /ɑ/ contrast, and the training effect persisted after six and twelve months (Escudero & Williams, 2015).

21.1.3 Perception and Production

A third area of longstanding interest in L2 phonological acquisition research is the extent to which L2 production modification reflects non-native perception and the related question of whether L2 perception and production develop in tandem. It is generally assumed that in children's L1 acquisition, accurate perception precedes accurate production (Smolensky, 1996); similarly for L2 acquisition, a number of researchers have suggested that many of the difficulties in L2 production stem from inaccurate perception of L2 targets.

As in so many areas, the literature provides conflicting evidence. Flege, Bohn, and Jang (1997) found correspondence between production and perception in the use of spectral and temporal cues for English vowel contrasts by learners from various language backgrounds. Yet other studies report no cross-modal correspondence or find one modality lagging behind the other. As discussed above, Hayes-Harb and Masuda (2008) found English-speaking learners of Japanese were able to perceive but not reliably produce the consonantal length contrast in novel words. On the other hand, Sheldon and Strange (1982) found that some Japanese learners of English were more successful in producing the English /l/ vs. /ɹ/ contrast than in discriminating these sounds.

The link between perception and production and the direction of influence between the two have been probed via studies in which learners are trained in one modality, and are tracked for improvement. The overall findings are that in many cases, but not all, training in either modality improves performance in the other, untrained, modality, with the magnitude of the effect dependent on various factors including the type of sound being trained (see Bradlow 2008; Rato & Oliveira, 2023; Sakai & Moorman, 2018 for reviews).

Recent work recognizes the dynamic nature of the perception-production link in L2 acquisition. Flege and Bohn (2021) propose that L2 production and perception "coevolve without precedence," in a departure from Flege (1995)'s hypothesis that accurate perception is a prerequisite of accurate production, pointing to the need for longitudinal studies to track the time-varying relationship between production and perception (Nagle & Baese-Berk, 2022), as well as for developing methodological consistencies in the tasks and statistical methods used to measure performance across two different modalities and across different studies (Schertz & Clare, 2020).

The remainder of this chapter will survey results of L2 acquisition research in the context of the questions outlined above, focusing on segmental acquisition (Section 21.2), phonotactics (Section 21.3), and prosody (Section 21.4).

21.2 Segmental Acquisition

The vastness of the literature on L2 segmental acquisition makes it impossible to provide a comprehensive review, so we will limit our discussion to some of the major research results on the production and perception of laryngeal contrasts in stop consonants and place contrasts in vowels.

21.2.1 Laryngeal Contrasts in Stop Consonants

Languages differ in their realization of laryngeal contrasts along the dimension of Voice Onset Time, or VOT (the time between the release of stop constriction and the onset of voicing on the following vowel). Even languages that share a two-way voicing or laryngeal contrast may implement this contrast differently (Beckman, Jessen, & Ringen, 2013). For example, in "aspirating" languages such as English, German, and Danish, /p t k/ have a long lag in voicing, resulting in aspiration, and /b d g/ have a short lag, so that voicing begins simultaneously with release of the stop into the following vowel. In “true voicing” languages such as Spanish, French, Dutch, Greek and Portuguese, however, /p t k/ have a short lag (are unaspirated), while for /b d g/ voicing begins during the stop closure (i.e. prevoiced). Many studies have found that when the L1 and L2 differ in phonetic realizations of voicing contrasts, L2 stops tend to show compromise VOT values intermediate between the L1 and the L2 stops (e.g., Fowler et al., 2008; see Zampini, 2008 for a review). For example, Flege (1987) found that the mean VOT duration of French /t/ for less experienced English-speaking learners of French was similar to the mean value for English /t/ produced by monolingual English speakers. For more experienced learners, however, the VOT was intermediate between the French norm and the English norm. These and similar findings are consistent with the claims discussed above that learners tend to map L2 sounds onto similar L1 phoneme categories, even when the L1 and L2 categories may differ in phonetic detail (Best, 1995; Flege, 1995; Kuhl & Iverson, 1995).

Another common finding is that L1 phonetic implementation patterns are not necessarily uniformly transferred. Learners whose L1 is a true-voicing language such as Spanish, French and Dutch tend to produce English voiceless aspirated stops as targetlike, i.e. with a long lag, but to produce English voiced stops as prevoiced, showing transfer from the L1 (Hazan & Boulakia, 1993; Simon, 2009; Williams, 1977). A similar asymmetry is found in English stops produced by heritage speakers of Tagalog, another true voicing language (e.g., Kang, George & Soo 2016). It is puzzling that the short lag stop category is not correctly produced in the L2 even though it is already available in the L1. Simon (2009) proposes that the privileged status of stops specified as [voice] makes them more likely to be transferred to L2 than voiceless stops, which lack such specification.

L1 transfer is found in the perception of L2 voicing contrasts as well as in production. Languages can differ in the relative importance of different acoustic cues (e.g, VOT vs. pitch cues on the following vowel) in distinguishing similar voicing contrasts, and learners adjust the relative cue weights in L2 acquisition. For example, Kong and Yoon (2013) found that compared to high-proficiency Korean learners of English, low-proficiency learners were more reliant on pitch, which is a primary cue for the Korean laryngeal contrast but a redundant secondary cue for the English voicing. In addition, Schertz et al. (2016) showed that Korean speakers were able to adjust their cue use in English voicing perception based on the distributional information in the training input.

21.2.2 Place Contrasts in Vowels

Both the perception and the production of L2 vowels are strongly influenced by listeners’ L1. For example, English speakers identified a French /y/ vowel as English /u/, while Portuguese speakers identified the same vowel as Portuguese /i/ (Rochet, 1995), equating the L2 vowel with an L1 category in perception, even though both English and Portuguese have /i/ and /u/. This perceptual distortion was also reflected in these speakers’ imitative production of French /y/.

As with laryngeal contrasts, L2 speakers may rely on acoustic cues to vowel contrasts that are different from those utilized by native speakers, and when several cues to a contrast are present, L1 and L2 speakers may weight them differently. Thus, many learners of English who have difficulty perceiving the English /i/ vs. /ɪ/ contrast may rely on a durational difference rather than vowel-quality differences, the primary cues for native speakers of most dialects. Such a pattern is found not only for speakers whose L1 has durational contrasts in vowels—German (Bohn, 1995) and Finnish (Ylinen et al., 2009)—but also for speakers whose L1 has no durational contrasts—Spanish, Mandarin, Korean, Russian and Catalan (Bohn, 1995; Cebrian, 2006; Escudero & Boersma, 2004; Flege et al., 1997; Morrison, 2009).

The reliance of L2 speakers on durational cues has been explained in terms of both L1 transfer and universal tendencies. Bohn (1995) proposes that when the vowel space containing two L2 vowels corresponds to a single vowel in L1, listeners will be "desensitized" to spectral distinctions within the region of the vowel space occupied by the L1 vowel, instead relying first on durational cues, which are allegedly more salient. However, Kondaurova and Francis (2008) suggest that the use of duration for L2 vowel contrasts by Spanish and Russian speakers can be analyzed as transfer, since durational cues may be used in the L1 as a secondary cue for distinctions in stress and consonant voicing. Arguments against the view that duration is a universally salient cue for L2 contrasts independent of L1 come from McAllister, Flege and Piske (2002) and Ylinen et al. (2005), who found that quantity distinctions in L2 vowels (Swedish and Finnish) are better perceived by speakers whose L1 uses duration to signal vowel contrasts (Estonian) than by speakers whose L1 does not contrast long and short vowels (Spanish, Russian). Relatedly, Chládková, Boersma, and Escudero (2022) found unsupervised distributional learning of novel vowel contrasts by Spanish speakers, where speakers are passively exposed to stimuli representing the target distribution of contrasting categories in a short training session, was successful when the new contrast involved a spectral boundary shift of an existing L1 contrast (/e/ vs. /i/), but not when it involved a novel durational contrast.

The acquisition of L2 vowel contrasts is complicated by the fact that individual vowel sounds may be realized differently in different contexts. Levy and Strange (2008) found that in the perception of /u/ vs. /y/ in French, inexperienced English-speaking learners showed more errors in the context of alveolar consonants, where English /u/ is allophonically fronted and therefore more similar to /y/, than in bilabial contexts, where it is not. Although the /u/ vs. /y/ contrast continued to be difficult for experienced speakers, the consonantal context effect disappeared, indicating that L2 learning includes all language-specific variations within phonetic categories. Levy (2009) also found that the English listeners' mapping of French front rounded vowels /y/ and /œ/ to English vowels showed less variation due to consonantal contexts for listeners with more L2 experience, indicating that the experienced learners' mapping applies at the phonemic rather than allophonic level. Similarly, in production, Oh (2008) examined the degree of coarticulation in alveolar stop+/u/ sequences in French and English speakers' L1 and L2, and found that English had more extensive C-to-V coarticulation while French had more extensive V-to-C coarticulation. Although many learners acquired both target vowels and coarticulation patterns, some acquired only target values, suggesting that coarticulation patterns are language-specific and must be learned independently of the target vowels.

21.3 Acquisition of Restrictions on Syllable Structure and of Phonotactics

In addition to differences in segment inventory, languages may differ in the number and types of segments that may be grouped into syllables and in the positions in which the segments may occur. In general, while research on the acquisition of L2 segments has tended to focus on the effects of L1 transfer, research on the acquisition of phonotactics and related phonological processes in the L2 sound system has been the source of a number of arguments for the role of universal markedness effects in L2 acquisition.

21.3.1 Consonant Clusters Within and Across Syllables

The perception and production of non-native consonant clusters has been extensively studied, in large part because this area provides evidence for emergent hierarchies independent of the L1 or L2 (Berent et al., 2008; Berent et al., 2009; Broselow & Finer, 1991; Davidson, 2010; Hancin-Bhatt, 2000; Hansen, 2004). Many of these studies address the question of whether ease of acquisition is related to sonority, assuming a universal sonority scale consisting of stops < fricatives < nasals < liquids < glides < vowels, organized from least to most sonorous. Two proposed universal principles govern the organization of segments within syllables: languages prefer sequences that increase in sonority approaching the vocalic nucleus, favoring rising sonority onsets such as stop-liquid over falling sonority onsets, and languages prefer clusters in which consonants are more distant in sonority over clusters containing consonants closer on the sonority scale. The specific requirements on sonority slope and distance in clusters can be parameterized by language, leading to cross-language differences.

Broselow and Finer (1991) found lower error rates for clusters with larger sonority distance such as /pr/ (stop–liquid) than for clusters with smaller sonority distance such as /fr/ (fricative–liquid) in the production of English pseudo-words by speakers of Japanese and Korean, where the only onset clusters allowed are obstruent-j. Similarly, Berent et al. (2008) and Berent et al. (2009) found that English and Korean speakers perceived non-native onset clusters of rising sonority more accurately than clusters of falling sonority. Relatedly, errors in production of English sC onsets by L1 Brazilian Portuguese speakers in a study by Cardoso (2008) correlated with sonority markedness, even though the marked onsets were more frequent in the input to the learners. Yet the sonority markedness effect has alternatively been attributed to statistical properties of featural combinations in the L1 lexicon (Albright, 2009) and to general phonetic properties; in a study of English and Catalan native speakers' productions of pseudo-words containing non-native word-initial clusters, Davidson (2010) argued that accuracy was better predicted by general phonetic factors such as articulatory ease and perceptual salience than by phonological markedness.

Native language restrictions also affect the perception of L2 structures. For example, English "pub" is borrowed into Japanese as *pabu* with a vowel inserted after /b/ because singleton /b/ cannot occur in Japanese syllable codas. Dupoux et al. (1999) investigated the perception of illegal structures such as the pseudo-word *ebzo*. Presented with a series of stimuli ranging from, for example, *ebzo* to *ebuzo*, with a vocalic portion of varying length (from null to a full vowel), the majority of the Japanese speakers perceived an "illusory vowel": they perceived *ebzo* as the possible native language structure *ebuzo* even when there was no vocalic signal in the stimulus. In contrast, the majority of French speakers, for whom *ebzo* is a possible structure, perceived a vowel between /b/ and /z/ only when the vocalic portion was at least 50ms long. However, discrimination of CC–CvC can improve when the difference is associated with lexical contrasts; Davidson, Shaw, and Adams (2007) found that American English speakers asked to distinguish pseudo-words containing a non-native CC sequence vs. a CvC sequence in picture-naming tasks performed better when the stimuli were presented with contrastive meanings.

Proponents of the view that illusory vowels arise specifically in response to native language constraints on possible syllables include Matthews and Brown (2004), who found that Thai speakers could discriminate heterosyllabic /k.t/ vs. /kVt/ better than Japanese speakers, even though the sequence /kt/ is not possible in either language, presumably because /k/ is a possible syllable coda in Thai (when not followed by /t/) but not in Japanese. Similarly, Kabak and Idsardi (2007) argued that Korean listeners perceived an illusory vowel only when the foreign consonant sequence includes an impossible coda in L1. Others have argued, however, that the asymmetries in vowel epenthesis across different languages and cluster types can be explained by factors including both top-down phonological knowledge and bottom-up acoustic information (Boersma & Hamann, 2009; Dupoux et al., 2011; Wilson, Davidson, & Martin, 2015; Daland, Oh, & Davidson, 2019; Whang, 2021).

Illusory vowel perception also evolves with L2 experience. Nomura and Ishikawa (2018) found that L1 Japanese intermediate-level English learners perceived an illusory vowel (e.g., reporting Japanese mora /mu/ in *homesick*) less often than introductory-level learners; they proposed that higher-proficiency learners can switch their attention to L2 phonotactics more readily, inhibiting L1-driven perception. Similarly, Kwon (2017) found that early Korean-English bilinguals' perception of an illusory vowel following a word-final stop was not affected by the presence or absence of stop release, while Korean learners of English were more likely to perceive an extra vowel when the stimuli had an audible release burst, suggesting that the two groups accessed different levels of L2 representation.

21.3.2 Restrictions on Segments in the Syllable Coda

Many languages place restrictions on the types of consonant contrasts that can be realized in syllable-final (coda) position. One common pattern is a prohibition on voiced obstruents in the syllable coda (as in Dutch, German, and Russian), resulting in a lack of obstruent voicing contrasts in that position. In an investigation of German speakers' L2 production of the English voicing contrast in word-final stops, Smith, Hayes-Harb, Bruss, and Harker (2009) found that the German speakers showed partial acquisition of the English contrast but their English production still fell short of L1 English speakers' performance.

More interesting cases involve emergent difficulty hierarchies that are not straightforwardly accounted for by either L1 transfer or L2 input. One such hierarchy emerges in Cardoso's (2007) examination of Brazilian Portuguese speakers' production of English word-final stops, in which L2 coronal stops were more frequently produced than stops in other places of articulation. Although Brazilian Portuguese allows no stops at all in coda, L2 coronal stops are more frequently produced than stops at other places of articulation, in line with the observation that crosslinguistically, coronal is the least marked of the three major places of articulation (Paradis & Prunet, 1991). Similarly, even learners whose L1 permits no final voiced obstruents are frequently more successful in producing L2 voiceless before voiced obstruents in word-final position, a pattern in accord with the ban on final voiced obstruents in numerous languages (e.g., Dutch, Russian, Catalan). In a review of the acquisition of final laryngeal contrasts in various L1-L2 pairings, Broselow (2018) discusses both formal analyses in terms of a putatively universal markedness constraint of the phonological grammar which penalizes voiced obstruents in syllable coda and research demonstrating the articulatory difficulty and lack of perceptual salience of obstruent voicing in coda position.

The hierarchies of difficulty discussed above are horizontal, revealing different rates of acquisition for different members of a class (e.g., consonant clusters). In addition, James (1987) suggests a vertical hierarchy whereby the lexical level of L2 phonological representation (e.g., phonemes and word-level accents) is acquired before higher levels of representation (i.e. phrasal properties). We turn now to the acquisition of prosodic systems as they are realized at the word, phrase and sentence levels.

21.4 Prosodic Systems: Stress, Pitch Accent, Tone and Intonation

Like the acquisition of L2 segments and phonotactics, the acquisition of prosody shows evidence for effects of the L1 as well as patterns not obviously grounded in either the L1 or the L2 input. Cross-language similarity in the function of prosodic cues may confer an advantage in learning a new prosodic system; for example, Altmann (2006) found that native speakers of Arabic (a stress language) were more successful in producing native-like English stress than were native speakers of Mandarin (a tone language), even though the placement of stress differs in Arabic and English, and So and Best (2010) found that native speakers of Hong Kong Cantonese (a tone language) and Standard Japanese (a pitch accent language) outperformed English speakers in identifying the tones of Mandarin syllables. At the same time, cross-language differences in the function of an acoustic cue may disadvantage learners: speakers of English, in which the major function of pitch is to encode discourse-level information such as sentence type, information structure, and speaker beliefs or attitudes, may have difficulty using pitch to signal lexical contrasts in Mandarin (Wang, Jongman, & Sereno, 2006), and in Japanese (Muradás-Taylor, 2022) while conversely, speakers of Mandarin may exhibit interference in their recognition of English words from variation in pitch in question versus statement intonation (Ortega-Llebaria, Nemogá, & Presson, 2015). Another major challenge in L2 prosody arises from the centrality of different acoustic cues in different prosodic systems. For example, native speakers of English rely mainly on vowel quality, duration, and intensity to encode word stress, while L1 speakers of Mandarin and Japanese, for which pitch is the main signal of lexical contrasts, may exhibit over-reliance on pitch cues to signal stress in English (Aoyama & Guion, 2007; Zhang, Nissen, & Francis, 2008). Below we discuss the acquisition of different types of word-level prosody (stress, tone and pitch accent) as well as the acquisition of intonational melodies superimposed on phrases and sentences.

21.4.1 Acquisition of Word Stress

While stress languages share the characteristic that syllables within a word differ in metrical prominence, languages may differ in whether the position of stress is phonologically predictable, lexically specified, affected by morphological structure, or some combination of these factors (as in English) and in the acoustic cues distinguishing stressed and unstressed syllables. The acquisition of a stress-based L2 may therefore pose difficulties even for learners whose L1 is stress-based.

No clear consensus exists on the time course of stress acquisition, the level of attainment that learners can be expected to reach, the relative difficulty of particular aspects of L2 stress. Studies by Guion (2005), Guion, Harada, and Clark (2004), and Wayland et al. (2006) have found systematic differences between early and late learners of English; a major determinant of L2 stress placement was the phonological similarity of the pseudo-word to existing English words, and surprisingly, “late learners of English may rely more heavily on word-by-word learning of stress patterns and are less likely to abstract generalities about stress placement” (Wayland et al., 2006, p. 298). In a survey of L2 stress research, van der Pas and Zonneveld (2004, p. 125) claim that “despite more optimistic claims, the bulk of this literature fails to demonstrate that [parameter] resetting may occur” though Özçelik (2021) does present evidence for at least partial parameter resetting, resulting in a grammar intermediate between L1 and L2.

Since learners may vary in their mastery of the production and the perception of L2 stress, we will consider these areas separately. Many studies of L2 stress production support the influence of the L1 system. The most common error in Archibald's (1993, 1994) studies of speakers of L1 Polish and Hungarian, where stress occurs in a fixed position in the word, involved placement of L2 stress in the L1 position. Youssef and Mazurkewich (1998) found that speakers of Cairene Arabic, where the position of stress varies depending on syllable weight, produced near-perfect placement for English words consistent with the Arabic pattern of stress on a superheavy (CVCC or CVVC) final syllable (*volunTEER*) or on a heavy (CVC or CVV) penultimate syllable (*moMENtum*), while words whose stress deviated from the L1 pattern were produced correctly less than half the time, with the vast majority of errors involving putting stress in the appropriate L1 position (*bariTONE*, *cyLINder*). In Kijak's (2009) study of L2 Polish (with fixed penultimate stress) by speakers of both fixed stress and variable stress L1s, the majority of non-L2 stresses involved stress on a possible L1 stress position, and learners whose L1 included penultimate stress as a possible option (English, German, Italian, Russian and Spanish) showed an initial advantage over those whose L1 allowed only final or initial stress. At the same time, speakers of Mandarin, who presumably did not experience interference from a native language stress system, were relatively successful in their mastery of Polish stress; in this case, as with segmental acquisition, the distance between the L1 and L2 actually seemed to facilitate learning by preventing equivalence classification based on similarities.

Another area of L1 influence in production involves the choice of acoustic cues used to convey stress; for example, Face (2003) found that the durational difference between stressed and unstressed vowels in Spanish was considerably larger for L1 English speakers living in Spain than for native speakers of Spanish, consistent with transfer from English patterns. L1 influence also appears in the reduction of unstressed syllables associated with languages such as English and German. English vowel reduction has been shown to be problematic for speakers of Spanish (Flege & Bohn, 1989), Japanese (Aoyama & Guion, 2007), and Mandarin (Zhang, Nissen & Francis, 2008), and a study of the L2 German of Korean and Spanish speakers (Young-Scholten, 1993) revealed that Spanish and Korean speakers learning German had difficulty in producing L2 inflectional affixes containing reduced vowels.

While L1 influence is clearly a factor in L2 production, some patterns have emerged that appear independent of both the L1 and L2, but that nonetheless conform to patterns found in human languages. Studies of the acquisition of English stress by speakers of Hungarian, Polish and Spanish (Archibald, 1993, 1994, 1997) as well as acquisition of stress in Khalkha Mongolian by speakers of English (Özçelik, 2021) have been argued to present cases of this type. One factor that emerges in several studies, in both production and perception, is a preference for assigning stress to heavy (CVV or CVC) as opposed to light (CV) syllables, and in particular to syllables containing long vowels. Kijak (2009) found that learners of L2 Polish tended to prefer stress on closed syllables in Polish, even when their L1 was also quantity-insensitive. Similarly, Guion et al. (2004) and Guion (2005) found a preference for placing stress on long vowels in studies of English L2 produced by speakers of Spanish and of Seoul Korean, a pattern that is not directly motivated by either Korean or Spanish, although a corpus study of English reveals that long vowels are roughly twice as likely to be stressed as are short vowels.

A related finding from crosslinguistic studies of native language systems is a tendency for vowels of higher sonority (i.e. vowels that are lower in height) to attract stress (De Lacy, 2006), presumably since higher sonority correlates with greater inherent duration. The existence of a natural bias toward this stress–sonority link was supported by the finding that both English and French speakers had more difficulty learning an artificial language in which lower-sonority vowels attracted stress than one in which stress fell preferably on vowels of greater sonority (Carpenter, 2010), even though neither English nor French appears to provide direct support for such a pattern. The stress–sonority correlation is also attested in regionalized varieties of English; Peng and Ann (2001) report a correlation between inherent vowel length and stress in Singapore English and Nigerian English, as well as the English of native Spanish speakers, with pronunciations like *CHInes*e, where /aj/ of the first syllable attracts stress in preference to the inherently shorter /i/ of the second syllable. The preferences for stressing heavier syllables and more sonorous vowels are therefore strong candidates for universal biases that may play a role in both language acquisition.

To summarize, the body of research on the production of L2 stress suggests a correlation between the similarity of the L1 and L2 prosodic systems and success in the acquisition of L2 prosody, as well as a tendency for L2 stress errors to reflect the native language stress system. Studies also support a role for universal biases favoring the placement of stress on heavier over lighter syllables and on longer over shorter vowels, which appear to emerge even in the absence of direct evidence from the L1 or L2.

Some of the studies discussed above argue for transfer in the perception of L2 stress as well as in production. Two major sources of difficulty in perception have been proposed: lack of attention to stress in the L2 input (so-called stress deafness) and misinterpretation of the L2 acoustic cues to stress.

Arguments for stress deafness come from a set of native language studies in which speakers of languages with fixed, predictable stress (Finnish, French and Hungarian) exhibited higher error rates and slower reaction times on tasks involving discrimination of stress differences than on tasks involving discrimination of phoneme differences. In contrast, speakers of Spanish, in whose native language stress is to some extent lexically determined, were equally sensitive to stress and segmental differences (Dupoux, Peperkamp & Sebastián-Gallés, 2001; Peperkamp, 2004; Peperkamp & Dupoux, 2002). The authors hypothesized that speakers of a language with fully predictable stress do not store stress in their phonological representations, although based on a study of Polish speakers' perception of stress, Peperkamp, Vendelin, and Dupoux (2010) concluded that stress deafness is gradient, correlating with the degree of predictability of stress in the speaker's language. Nonetheless, in Kijak's (2009) study, speakers of Czech, which has regular word-initial stress, performed very well in the perception of Polish stress; as Kijak points out, fixed stress on a particular position in a word may help in segmenting the continuous speech stream into words, providing Czech speakers with good reason to attend to stress in their L1. The importance of prosodic cues in the segmentation of continuous speech into words is further supported by an eye-tracking study (Tremblay, Broersma, & Coughlin, 2018) comparing the ability of L1 Dutch and L1 English speakers to recognize French words in different positions in French sentences. The rising F0 associated with phrase-final position in French was more useful in word recognition for the Dutch speakers than for the English speakers, presumably because pitch cues play a greater role in signaling stress in Dutch than in English, where vowel reduction is the most salient L1 stress cue. Similarly, Altenberg (2005) found that Spanish-speaking learners of English were significantly worse than native English speakers at using cues such as aspiration, which are irrelevant in their L1, in distinguishing phrases such as *keep s[p]arking* and *keeps [p^h]arking*.

If a pattern of fully predictable word stress in L1 might interfere with learners' ability to perceive L2 stress, what about the absence of stress in the native language? Guion (2005), for example, found that Korean–English bilinguals were less like English L1 speakers in their perception of English stress than were Spanish–English bilinguals, and Nguyen, Ingram and Pensalfini (2008) found that speakers of Vietnamese, a tone language, had difficulty in distinguishing compound nouns with initial prominence from noun phrases with prominence on the final head (e.g., *BLACKberry* versus *black BERRY*). However, speakers of an L1 stress language are not necessarily advantaged in perceiving L2 stress. In a study by Ortega-Llebaria, Gu and Fan (2013), L1 English speakers had difficulty in identifying whether the *mama* in a Spanish sentence bore initial or final stress, due to differences in the use of acoustic cues to stress in the two languages. Numerous studies (e.g., Chrabasz et al., 2014; Connell et al., 2018, Nguyen, Ingram and Pensalfini, 2008); Chen (2007) have supported the claim that the L1 system may encourage listeners to interpret L2 acoustic cues through the filter of the L1 phonology; for example, Chen (2007) found that Mandarin-speaking learners of Spanish tended to interpret penultimate stress in Spanish words as a lexical rising tone on the penult followed by a falling tone on the following syllable. Some studies suggest that non-native speakers may actually enjoy an advantage over native speakers in L2 perception tasks involving L2 prosodic cues that serve some important (though different) function in the L1. Speakers of the pitch accent language Gyeongsang Korean outperformed native English speakers in sequence recall of English words differing in stress position (Kim & Tremblay, 2021); L1 Mandarin learners of Japanese showed higher accuracy than L1 Japanese speakers in discriminating Japanese forms with different pitch accent patterns (Wiener & Goss, 2019); and speakers of Cantonese, a tone language, outperformed English speakers in AX discrimination of English natural stimuli differing in stress position, though this advantage disappeared when stimuli were manipulated to remove pitch cues (Choi, Tong & Samuel, 2019; Choi, 2021).

21.4.2 Acquisition of Tone, Pitch Accent and Intonation

Studies of cross-language tone perception suggest that an L1 system in which pitch is used to realize lexical contrasts confers some advantage in tone perception; for example, Lee, Vakoch and Wurm (1996) found that native speakers of Cantonese performed better than English speakers in identifying pairs of Mandarin syllables as same or different in tone, even though the Cantonese tone inventory is different from that of Mandarin. Similarly, though L1 Cantonese speakers showed an advantage over L1 English speakers in identification of Mandarin tones, their L1 knowledge actually proved a disadvantage in one respect; the Mandarin tones that they found most difficult to discriminate were those that fell within the range of a single tone category in Cantonese, a pattern reminiscent of equivalence classification in the perception of phonemes (So & Best, 2010).

Another type of equivalence classification involves interpreting pitch cues used for lexical contrasts in the L2 in terms of the intonational patterns of the L1. Broselow, Hurtig and Ringen (1987) found that for English speakers who received training in identifying the four Mandarin tones in sequences of syllables, the most frequent error involved (mis)identification of Tone 4 (falling tone) in final position as the high level Tone 1, presumably reflecting interpretation of the falling pitch of Tone 4 as the fall associated with the end of declarative sentences, with only the high starting point of the syllable interpreted as lexical tone.

Conversely, speakers of an L1 in which pitch signals lexical contrasts appear to have difficulty using L2 pitch cues associated with discourse-level focus in English. When Wayland et al. (2019) presented L1 Mandarin speakers with English SVO echo questions spoken with prosody appropriate for focus on either the verb or the object, the Mandarin speakers were significantly less accurate than the L1 English speakers in choosing the most appropriate response. In a study of the production of English focus prosody, Graham and Post (2018) found that L1 Spanish speakers produced more English-like contours than L1 Japanese speakers, though both groups differed from native speakers of English in their alignment of the pitch peak with the focused item. And while native speakers of English have been shown to use contrastive prosody to anticipate upcoming information, Mandarin speakers have not shown the same ability. English listeners in Ito and Speer's (2008) eye tracking experiment who heard instructions with contrastive prosody on an adjective (*Hang the green ball; Next, hang the BLUE ball*) looked toward the blue ball, rather than another blue object, even before hearing the relevant noun. Yet similar studies comparing English and Mandarin listeners (Lee, Perdomo & Kaan, 2020; Perdomo & Kaan, 2021) suggest that L1 Mandarin listeners are less able than L1 English speakers to use English contrastive prosody in processing, despite the attentiveness to pitch demanded by their native language. The lack of tone language advantage appeared even in a study which forced listeners to rely on prosodic cues alone, by including stimuli in which the segmental information of the target noun phrase was masked, as in *Click on the scarlet mittens. Now click on the* ~~*PURple mittens*~~ (where purple mittens were the only prosodically appropriate item on the screen, and low-pass filtering made the segments of *purple mittens* unrecoverable while leaving the prosody intact). Here too the L1 Mandarin speakers performed worse in using contrastive prosody than the L1 English speakers, despite the importance of pitch in their L1 (Hwang et al., 2022). These results contrast with those discussed above in which L1 tone language speakers showed an advantage in discriminating both English stress patterns (Choi et al., 2019) and Japanese pitch accent patterns (Wiener & Goss, 2019). This contrast suggests that native language transfer confers an advantage when the functions of prosodic cues in the two languages overlap (e.g., encoding lexical contrasts in both the L1 and L2), but not when they differ (encoding lexical information in the L1 but discourse-level information in the L2).

Even where the L1 and L2 make similar use of pitch, the meanings of intonation contours differ in the two languages. In tasks designed to determine how English-speaking learners of Portuguese and Portuguese-speaking learners of English interpreted L2 intonational contours, Cruz-Ferreira (1987) found that both groups performed similarly where the same meaning was conveyed by similar intonational contours in L1 and L2, but had difficulty when the intonational contour conveyed different meanings. Similarly, Chen (2009) found interference from L1 prosody in the perception of emphasis, which is typically expressed in English by use of a lower pitch register but in Dutch by use of a higher pitch register in English,. In listening to English, L1 Dutch learners of English associated emphasis with higher pitch, interpreting the pitch cues in terms of their Dutch meaning, although, L1 English speakers hearing Dutch did not show L1 interference. Chen suggests that this asymmetry may be due to an asymmetry in classroom input: only the English learners' L2 instruction came from native speakers of the L2.

A second source of L1 influence may come from differences in the phonetic realization of pitch contours. A comparison of the production of English contradiction prosody by L1 Mandarin and native English speakers revealed that the English speakers tended to align the high pitch with the stressed syllable of the focused word, while the L1 Mandarin speakers were more likely to produce the entire focused word with higher pitch (Kao et al., 2016), and Liu and Reed (2021) show differences between L1 Mandarin and L1 English speakers in producing and processing various acoustic cues to sentence stress. Bi-directional influence emerged in a study of speakers whose two languages have relatively similar intonational systems (Mennen 2004). In both Dutch and Modern Greek, declarative intonation includes a (non-final) rise (LH*), but in Dutch, the peak occurs on the accented syllable, while in Greek, the peak occurs following the accented syllable. Four out of their five L1 Dutch learners of Greek produced peak alignment patterns that were different from those of both native control groups, not only in their L2 Greek, but also in their L1 Dutch, showing that the interlanguage influence may go in both directions.

Another area where L1 influence may be found is in the use of prosodic cues to disambiguate phrases or sentences that allow for multiple interpretations. Yang (2010) found that while L1 English speakers relied on pre-boundary lengthening and pause to disambiguate the scope of the adjective in phrases like *The little cats and dogs*, less proficient English-learning speakers of Taiwanese did not produce reliable differences. In studies comparing both the production and the perception of ambiguous English structures such as *The boss of the clerk who was dishonest,* along with the production of comparable ambiguous Korean structures, Baek (2021, 2022) found that L1 Korean speakers differed from L1 English speakers in relying mainly on pause and intensity cues to distinguish the readings in which the relative clause modified *the boss* or *the clerk*, paralleling their reliance on these cues in disambiguating comparable structures in Korean.

21.5 Conclusion

Much research in L2 production, perception, and processing of speech has focused on whether L2 acquisition patterns should be understood as effects of L1 transfer or of universal preferences for particular linguistic structures; whether acquisition patterns should be explained at the level of abstract phonology or at a phonetic level; and whether patterns in L2 production correlate with patterns in L2 perception. Our goal in this chapter was to demonstrate that although these questions have provided a productive program for L2 research, the dichotomies they presuppose are overly simplistic. The literature has provided plausible examples both of transfer from the native language and of emergent patterns that appear not to be motivated by input from either L1 or L2. Similarly, some L2 patterns seem to be well explained with reference to abstract phonological structures while others can only be explained as effects of phonetic salience and/or articulatory ease; furthermore, as the boundary between phonetics and phonology has become increasingly blurred by work assuming that the constraints of the phonological grammar are grounded in phonetics (e.g., Hayes, Steriade, & Kirchner, 2004); that the phonological grammar may make direct reference to the timing of articulatory gestures (e.g., Browman & Goldstein 1989 and subsequent work in Articulatory Phonology); and that the output of the phonological grammar may be probabilistic rather than categorical (Boersma & Hayes, 2001; Hayes & Wilson, 2008). Nor has a simple explanation of the relationship between production, perception and lexical representations emerged, with accurate perception appearing to lag behind production in some areas and to precede it in others. Clearly, L2 speech represents a complex interplay of numerous factors, and despite many advances, no model has yet emerged to provide a fully comprehensive and predictive account of the patterns found in segmental and prosodic L2 speech (Nagle & Baese-Berk, 2022).

It is clear that much work in L2 phonology and speech remains to be done. There is a regrettable dearth of studies charting longitudinal development, particularly the development of L2 prosody. Sorely needed are more studies of L2 stress that take languages other than English as their L2, studies that investigate production and perception simultaneously with consistent methods, and studies that tap into both auditory and lexical levels. However, there are grounds for considerable optimism, as methodological advances in the field of language study transform the way research into L2 sound acquisition is conducted (Schmid & Dusseldorp, 2010). Behavioral probes such as eye-tracking methods and neurolinguistic tools, in particular ERP studies, provide fine-grained temporal information concerning the time course of L2 speech processing (e.g., Perdomo & Kaan, 2021; Chládková et al., 2022). The availability of ultrasound imaging and electropalatography allow direct observation of articulatory patterns (e.g., Mennen et al., 2010), and phonological theory now offers tools for modeling the fine-grained timing differences in L1 and L2 production (e.g., Zsiga, 2003). And while the role of input in L2 acquisition has always been recognized (e.g., Young-Scholten, 1994), increasingly large databases and increasingly sophisticated search tools may make possible the discovery of statistical tendencies in L2 input that might not otherwise be apparent (e.g., Guion et al., 2003), and developments in modeling statistically driven acquisition (Albright, 2009; Boersma & Hayes, 2001; Hayes & Wilson, 2008; Shaw & Davidson, 2011; Wilson & Davidson, 2013) may shed light on the extent to which L2 patterns which are seemingly unmotivated by L1 transfer or L2 input may be accounted for by the L2 learners' probabilistic knowledge of the learning data. Furthermore, by allowing experimenters to strictly control access to learning input, artificial grammar learning experiments (e.g., Carpenter, 2010) provide a tool to determine whether typological tendencies reflect genuine learning biases or simply a function of linguistic change (e.g., Blevins, 2004). Finally, explicit models have been offered to describe the mapping from the acoustic signal to phonological representations (e.g., Escudero & Boersma, 2004; Escudero, 2009), an area long neglected in formal theories of phonology. Inevitably, new paradigms will emerge in this swiftly evolving field to offer fresh perspectives on the perennially intriguing questions posed by second language speech patterns.

References

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology, 26*, 9-41.

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

Altmann, H. (2006). *The perception and production of second language stress: a cross-linguistic experimental study* [Doctoral dissertation, University of Delaware]. ProQuest Dissertations & Theses Global.

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

Aoyama, K. & Guion, S. (2007). Prosody in second language acquisition: An acoustic analysis on duration and F0 range. In O.-S. Bohn & M. J. Munro (Eds.), *The role of language experience in second-language speech learning*: *In honor of James Emil Flege* (pp. 281-297). John Benjamins.

Archibald, J. (1993). *Language learnability and L2 phonology: The acquisition of metrical parameters*. Kluwer.

Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research, 10,* 215-240.

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: lexical storage versus computation of stress. *Linguistics, 35*, 167-181.

Archibald, J. (2005). Second language phonology as redeployment of NL phonological knowledge. *Canadian Journal of Linguistics, 50,* 285-314.

Archibald, J. (2021). Ease and difficulty in L2 phonology: a mini-review. *Frontiers in Communication*, *6*, 626529.

Baek, H. (2021). Prosodic disambiguation in first and second language production: English and Korean. *Language and Speech, 65,* 598-624.

Baek, H. (2022). Prosodic cue weighting in the processing of ambiguous sentences by native English listeners and Korean listeners of English. *The Journal of the Acoustical Society of America, 151,* 158-167.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages1. *Journal of Linguistics*, *49*(2), 259-284.

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Science, 105*, 5321-5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: evidence from nasal clusters. *Phonology, 26,* 75-108.

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 speech research* (pp. 171–204). York Press.

Best, C. T., & Hallé, P. A. (2010). Perception of initial obstruent voicing is influenced by gestural organization. *Journal of Phonetics*, *38*(1), 109-126.

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

Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp.206-244). Blackwell.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Boersma, P. & Hamann S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & W. L. Wetzels (Eds.), *Loan Phonology* (pp. 11-58). John Benjamins.

Boersma, P. & Hayes, B. (2001). Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32,* 45-86.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 279–304). York Press.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast. In J. G. H. Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). John Benjamins.

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

Broersma, M. (2012). Lexical representation of perceptually difficult second-language words. *The Journal of the Acoustical Society of America*, *132*(3), 2053-2053.

Broselow, E. (2018). Laryngeal contrasts in second language phonology. In L. Hyman & F. Plank (Eds.), *Phonological Typology,* (pp. 312–340). Walter de Gruyter.

Broselow, E. & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research, 7,* 35-59.

Broselow, E., Hurtig, R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S. Weinberger (Eds.), *Interlanguage Phonology* (pp. 350–362). Newbury House.

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

Brown, C. (2000). Speech perception and phonological acquisition. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Blackwell.

Bundgaard-Nielsen, R.L., Best, C.T., & Tyler, M.D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Sanguage Acquisition*, *33*(3), 433–461.
Cardoso, W. (2007). The variable development of English word-final stops by Brazilian Portuguese speakers: A stochastic Optimality theoretic account. *Language Variation and Change, 19,* 219–248.
Cardoso, W. (2008). The development of sC onset clusters in interlanguage: Markedness vs. frequency effects. *Proceedings of the 9th Generative approaches to second language acquisition conference,* 15–29
Carpenter, A. (2010). A naturalness bias in learning stress. *Phonology, 27,* 345-392.
Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34,* 372–387.
Chang, C. B. (2015). Determining Cross-Linguistic Phonological Similarity Between Segments: The Primacy of Abstract Aspects of Similarity. In E. Raimy & C. E. Cairns (Eds.), *The segment in phonetics and phonology* (pp. 199–217). Wiley.
Chang, C. B. (2018). Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics, 68*, 85–102.
Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning, 59,* 367–409.
Chen, Y. (2007). From tone to accent: the tonal transfer strategy for Chinese L2 learners of Spanish. *Proceedings of ICPhS XVI*, 1645-1648.
Chládková, K., Boersma, P., & Escudero, P. (2022). Unattended distributional training can shift phoneme boundaries. *Bilingualism: Language and Cognition, 25*(5), 827-840.
Choi, W. (2021). Cantonese advantage on English stress perception: Constraints and neural underpinnings. *Neuropsychologia, 158*, 1-12.
Choi, W., Tonog, X., & Samuel, A.G. (2019). Better than native: Tone language experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners. *Cognition, 18*, 188-192.
Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research, 57*, 1468-1479.
Colantoni, L. & Steele, J. (2007). Acquiring /ʁ/ in context. *Studies in Second Language Acquisition, 29,* 381-406.
Connell, K., Hüls, S., Martínez-García,M., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (2018). English Learners' Use of Segmental and Suprasegmental Cues to Stress in Lexical Access: An Eye-Tracking Study. *Language Learning, 68*, 653-668.
Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 102-120). Foris Publications.
Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34,* 269-284.
Darcy, I., Dekydtspotter, L., Sprouse, R.A., Glover, J., Kaden, C., McGuire, M. & Scott, J.H., (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second Language Research*, *28*(1), 5-40.
Darcy, I., Mora, J.C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, *66*(4), 741-773.

Daland, R., Oh, M. & Davidson, L. (2019). On the relation between speech perception and loanword adaptation. *Natural Language and Linguistic Theory, 37,* 825-868.

Davidson, L. (2006). Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. *Cognitive Science*, *30*(5), 837-862.

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

Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass, 5,* 126-139.

Davidson, L., Shaw, J., & Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *Journal of the Acoustical Society of America, 122,* 3697-3709.

Daidone, D., & Darcy, I. (2021). Vocabulary size is a key factor in predicting second language lexical encoding accuracy. *Frontiers in Psychology*, *12*, 688356.

de Jong, K. J., Hao, Y.-C., & Park, H. (2009). Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics, 37,* 357–373.

de Jong, K. J., Silbert, N., & Park, H. (2009). Segmental generalization in second language segment identification. *Language Learning, 59,* 1–31.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1568–1578.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*(3), 199-210.

Dupoux, E., Peperkamp, S., & Sebastian-Galles, N. (2001). A robust method to study stress 'deafness.' *Journal of the Acoustical Society of America, 110,* 1606-1618.

Eckman, F. R. (1977). Markedness and the Contrastive Analysis Hypothesis. *Language Learning, 27,* 315-330.

Eckman, F. R. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition, 26,* 513-549.

Eckman, F. R., Iverson, G. K., Fox, R. A., Jacewicz, E. & Lee, S. (2003). Perception and production in the acquisition of L2 phonemic contrasts. *Second Language Research, 19,* 169-208.

Escudero, P. (2009). The linguistic perception of SIMILAR L2 sounds. In P. Boersma & S. Hamann (Eds.) *Phonology in Perception* (pp. 151-190). De Gruyter Mouton.

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

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetrical lexical access. *Journal of Phonetics, 36*, 345-360.

Escudero, P. & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition, 132*(2), 408-413.

Face, T. L. (2023). Production of acoustic correlates of stress by L2 Spanish-speaking immigrants to Spain. *Languages, 8,* 258.

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. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 233–277). York Press.

Flege, J. & Bohn, O.-S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies of Second Language Acquisition, 11,* 35-62.

Flege, J., & Bohn, O-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3-83). Cambridge University Press.

Flege, J., Bohn, O-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, *25,* 437-470.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. (2008). Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics, 36,* 649-663.

Goad, H. & White, L. (2006). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research, 22*, 243-68.

Graham, C. & Post., B. (2018). Second language acquisition of intonation: peak alignment in American English. *Journal of Phonetics, 66,* 1-14.

Guion, S. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition, 27,* 503-533.

Guion, S., Harada, T., & Clark, J.J. (2004). Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition, 7*, 207-26.

Hancin-Bhatt, B. (1994). Segmental transfer: A consequence of a dynamic system. *Second Language Research, 10*, 241-69.

Hancin-Bhatt, B. (2000). Optimality in second language phonology: codas in Thai ESL. *Second Language Research, 16*, 201-32.

Hayes, B., Steriade, D., & Kirchner, R. (Eds.). 2004. *Phonetically based phonology.* Cambridge University Press.

Hayes, B. & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 29,* 379-440.

Hayes-Harb, Rachel. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research, 23,* 65-94.

Hayes-Harb, R. & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research, 24,* 5-33.

Hazan, V. & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech, 36,* 17-38.

Hwang, J., Takahashi, C., Baek, H. Yeung, A., & Broselow, E. (2022). Do L1 tone language speakers enjoy a perceptual advantage in processing English contrastive prosody? *Bilingualism: Language and Cognition, 25,* 816-826.

Ingram, J. C. & Park, S. G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America, 103,* 1161-1174.

Ioup, G. (2008). Exploring the role of age in the acquisition of a second language phonology. In J. G. H. Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 41–62). John Benjamins.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonations: Eye movements during instructed visual search. *Journal of Memory and Language, 58,* 541-573.

James, A. (1987). Prosodic structure in phonological acquisition. *Second Language Research, 3,* 118-140.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus, 14,* 147-172.
Kabak, B., & Idsardi, W. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech, 50*, 23-52.
Kang, K.-H. & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America, 119,* 1672-1683.
Kang, Y. (2008). Interlanguage segmental mapping as evidence for the nature of lexical representation. *Language and Linguistics Compass, 2,* 103-118.
Kang, Y., George, S., & Soo, R. (2016). Cross-language influence in the stop voicing contrast in heritage Tagalog. *Heritage Language Journal*, *13*(2), 184-218.
Kao, S., Hwang, J., Baek, H., Takahashi, C., & Broselow, E. (2016). International teaching assistants' production of English focus marking. *Proceedings of Meetings on Acoustics, 26*(1).
Kijak, A. (2009). *How stressful is L2 stress? A cross-linguistic study of L2 perception and production of metrical systems*. Netherlands Graduate School of Linguistics.
Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics, 87,* 1-15.
Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America, 124,* 3959-3971.
Kong, E., & Yoon, I. (2013). L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: Production and perception of English stops. *Journal of the Korean Society of Speech Sciences*, *5*(4), 81–90.
Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 121–154). York Press.
Kwon, H. (2017). Language experience, speech perception and loanword adaptation: Variable adaptation of English word-final plosives into Korean. *Journal of Phonetics, 60,* 1–19.
Lado, R. (1957). *Linguistics across cultures.* University of Michigan Press.
Larson-Hall, J. (2004). Predicting perceptual success with segments: A test of Japanese speakers of Russian. *Second Language Research, 20,* 33-76.
Lee, A., Perdomo, M., & Kaan, E. (2020). Native and second-language processing of contrastive pitch accent: An ERP study. *Second Language Research, 36,* 503-527.
Lee, Y.-S, Vakoch, D.A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research, 25,* 527-542.
Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America, 125,* 1138–1152.
Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36,* 141-157.
Liu, D., & Reed, M. (2021). Exploring the complexity of the L2 intonation system: An acoustic and eye-tracking study. *Frontiers in Communication*, *6*, 627316.

Matthews, J. & Brown, C. (2004). When intake exceeds input: Language specific perceptual illusions induced by NL prosodic constraints. *International Journal of Bilingualism, 8,* 5-27.
Maye, J., Werker, J.F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.
McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of NL on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30,* 229-258.
Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics, 32,* 543-563.
Mennen, I., Scobbie, J. M., de Leeuw, E., Schaeffler, S., & Schaeffler, F. (2010). Measuring language-specific phonetic settings. *Second Language Research, 26,* 13-41.
Morrison, G. S. (2009). NL-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech, 52,* 437-462.
Muradás-Taylor, B. (2022). Accuracy and stability in English speakers' production of Japanese pitch accent. *Language and Speech, 65,* 377-403.
Nagle, C.L., & Baese-Berk, M.M. (2022). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, *44*(2), 580-605
Nguyen, T., Ingram, C., & Pensalfini, J. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics, 36,* 158-190.
Nomura, J., & Ishikawa, K. (2018). Effects of first language processes and representations on second language perception: The case of vowel epenthesis by Japanese speakers. *International Journal of Bilingualism, 22*(1), 69–87.
Oh, E. (2008). Coarticulation in non-native speakers of English and French: An acoustic study. *Journal of Phonetics, 36,* 361-384.
Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics, 41,* 186-197.
Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2015). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese-English bilinguals perceive "Rose?" vs. "Rose". *Bilingualism: Language and Cognition, 20*, 367-383.
Özçelik, O. (2021). L2 acquisition of a complex stress pattern: UG-constrained learning paths in Khalkha Mongolian. *Frontiers in Psychology, 12*, 627797.
Paradis, C., & Prunet, J.-F. (Eds.). (1991). *The special status of coronals: Internal and external evidence.* Academic Press.
Peng, L., & Ann, J. (2001). Stress and duration in three varieties of English. *World Englishes, 20,* 1-27.
Peperkamp, S. (2004). Lexical exception in stress systems: Arguments from early language acquisition and adult speech perception. *Language, 80,* 98-126.
Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 203-240). Mouton de Gruyter.
Peperkamp, S, Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics, 38,* 422-430.
Perdomo, M., & Kaan, E. (2021). Prosodic cues in second-language speech processing: A visual world eye-tracking study. *Second Language Research, 37,* 349-375.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29,* 191-215.

Prince, A. & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in Generative Grammar.* Blackwell Publishing.

Rato, A. & Oliveira, D. (2023). Assessing the robustness of L2 perceptual training: A closer look at generalization and retention of learning. In U. K. Alves & J. A. de Albuquerque (Eds.), *Second language pronunciation: Different approaches to teaching and training* (pp. 369-396). De Gruyter Mouton.

Rice, K. (2007). Markedness in phonology. In P. deLacy (Ed.), *The Cambridge handbook of phonology* (pp. 79-97). Cambridge University Press.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.) *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 379–410). York Press.

Sakai, M., & Moorman, C., (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, and Psychophysics*, *78*, 355-367.

Schertz, J., & Clare, E.J., (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(2), e1521.

Schmid, M. S., &and Dusseldorp, E. (2010). Innovative and quantitative methods for bilingualism research. *Second Language Research, 26,* 5-11.

Shaw, J., & Davidson, L. (2011). Perceptual similarity in input-output mappings: A computational/experimental study of non-native speech production. *Lingua, 121,* 1344-1358.

Sheldon, A., & 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,* 243–261.

Simon, E. (2009). Acquiring a new second language contrast: an analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research, 25,* 377-408.

Smith, B. L., Hayes-Harb, R., Bruss, M., & Harker A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics, 37,* 257-75.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry, 27,* 720–731.

So, C., & Best, C. (2010). Cross-language perception of Non-native tonal contrasts: effects of native phonological and phonetic influence. *Language and Speech, 53,* 273-293.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39,* 456–466.

Tremblay, A., Broersma, M., & Coughlin, C. E. (2018). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition*, *21*(3), 640-652.

Trofimovich, P., Gatbonton, E., & Segalowitz, N. (2007). A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition, 29,* 407-448.

van der Pas, B. & Zonneveld, W. (2004). L2 parameter setting for metrical systems. *The Linguistic Review, 21,* 125-170.

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

Wang, Y., Jongman, A., & Sereno, J. (2006). Second language acquisition and processing of Mandarin tone. In E. Bates, L. Tan, & O. Tzeng, *Handbook of East Asian psycholinguistics, volume 1: Chinese* (pp. 250-257). Cambridge University Press.

Wayland, R., Guion, S., Landfair, D., & Li, B. (2006). Native Thai speakers' acquisition of English word stress patterns. *Psycholinguistic Research, 35*, 285-304.

Wayland, R., Guerra, C., Chen, S., & Zhu, Y. (2019). English focus perception by Mandarin listeners. *Languages, 4,* 91.

Whang, J. (2021). Multiple sources of surprisal affect illusory vowel epenthesis. *Frontiers in Psychology, 12,* 677571.

Wiener, S., & Goss, A. (2019). Second and third language learners' sensitivity to Japanese pitch accent is additive: An information-based model of pitch perception. *Studies in Second Language Acquisition, 41,* 897-910.

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

Wilson, C. & Davidson, L. (2013). Bayesian analysis of non-native cluster production. *Proceedings of the 40th annual meeting of the North East Linguistic Society*, 265-278.

Wilson, C., Davidson, L., & Martin, S. (2015). Effects of acoustic detail on cross-language speech production. *Journal of Memory and Language, 77*, 1-24.

Wissing, D. & Zonneveld, W. (1996). Final devoicing as a robust phenomenon in second language acquisition: Tswana, English, and Afrikaans. *South African Journal of Linguistics, 34,* 3-24.

Yang, P.-L. (2010). English language proficiency and production of prosodic disambiguation: A preliminary study of Taiwanese English learners. *English Language and Teaching, 34*, 51-84.

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2023). Feature-driven new sound category formation: Computational implementation with the L2LP model and beyond. *Frontiers in Language Sciences*, *2*, 1303511.

Ylinen, S., Shestakova, A., Alku, P., & Huotilainen, M. (2005). The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish. *Language and Speech, 48,* 313-338.

Ylinen, S., Uther, M. Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2009). Training the brain to weight speech cues differently: A study of Finnish Second-language users of English. *Journal of Cognitive Neuroscience, 22,* 1319-1332.

Young-Scholten, M. (1993). *The acquisition of prosodic structure in a second language*. Max Niemeyer Verlag.

Young-Scholten, M. (1994). On positive evidence and ultimate attainment in L2 phonology. *Second Language Research, 10,* 193-214.

Young-Scholten, M. (2004). Prosodic constraints on allophonic distribution in adult L2 acquisition. *International Journal of Bilingualism, 8,* 67-77.

Youssef, A., & Mazurkewich, I. (1998). The acquisition of English metrical parameters and syllable structure by adult native speakers of Egyptian Arabic (Cairene dialect). In S. Flynn, G. Martohardjono, & W. O'Neil (Eds.), *The Generative study of language acquisition* (pp. 303-332). Lawrence Erlbaum.

Zampini, M. L. (2008). L2 speech production research: Findings, issues, and advances. In J. G. H. Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 219–249). John Benjamins.

Zhang, Y., Nissen, S., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America, 123,* 4498-4513.

Zsiga, E. C. (2003). Articulatory timing in a second language: Evidence from English and Russian. *Studies in Second Language Acquisition, 25,* 399-432.